UQAC
Université du Québec
à Chicoutimi

CONCEPTION ET IMPLÉMENTATION D'UN SYSTÈME INFORMATIQUE DE SURVEILLANCE EN TEMPS RÉEL POUR LA FABRICATION INTELLIGENTE APPLIQUÉE À LA SOUDURE DE L'ALUMINIUM

PAR OLIVIER BRIAND

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI EN VUE DE L'OBTENTION DU GRADE DE MAÎTRE ÈS SCIENCES (M. SC.) EN INFORMATIQUE

QUÉBEC, CANADA

RÉSUMÉ

Le procédé Gas Metal Arc Welding (GMAW) automatisé de l'aluminium est sensible à plusieurs dérives de procédé dont la détection rapide est essentielle pour éviter les pièces non conformes qui peuvent causer des pertes économiques. Ce mémoire présente la conception et l'implémentation d'un système de surveillance s'appuyant sur l'analyse de séries temporelles issues de signaux électriques (courant, tension) et acoustiques, afin de classifier en quasi-temps réel les anomalies qui peuvent survenir pendant la soudure.

Un jeu de données contrôlé, constitué en laboratoire, couvre les conditions anormales typiques (7 types d'anomalies) et une classe de référence. Les signaux sont prétraités et exploités à l'aide de deux modèles différents : un Random Forest (RF), algorithme de Machine Learning (ML) classique qui utilise des caractéristiques extraites automatiquement par TSFresh et un Convolutional Neural Network (CNN), basé sur le Deep Learning (DL). La comparaison des modèles et des sources de données met en évidence des complémentarités intéressantes : les RF sont plus efficaces avec des signaux électriques, tandis que les CNN exploitent mieux des signaux acoustiques. Les meilleures performances sont obtenues avec des modèles de RF avec données électriques. L'explicabilité des modèles de RF a été intégrée avec SHAP pour mieux interpréter les prédictions.

Une application web démontre l'intégration à la cellule de soudage. À la fin de chaque soudure, les prédictions et leur taux de confiance sont générées et affichés visuellement. Des essais après changements dans l'installation montrent que la généralisation reste limitée, mais qu'un réentraînement rapide est possible pour atteindre une classification binaire. Certaines pistes d'amélioration sont discutées dans le but de réduire les limitations actuelles et d'optimiser les performances du système.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

CBAM Convolutional Block Attention Module
CNN Convolutional Neural Network
CNRC Conseil National de Recherches Canada
CTA Centre des Technologies de l'Aluminium
DL Deep Learning
FCAW Flux-Cored Arc Welding
FFT Fast Fourier Transform
GMAW Gas Metal Arc Welding
IA Intelligence Artificielle
LOF Local Outlier Factor
LSTM Long Short-Term Memory
MC-DCNN Multi-Channels Deep Convolutional Neural Network
ML Machine Learning
PCA Principal Component Analysis
R&D Recherche et Développement
RF Random Forest
RNN Recurrent Neural Network
WAAM Wire Arc Additive Manufacturing
χ^2 Test du khi carré

REMERCIEMENTS

Ce projet a été réalisé dans le cadre des activités du groupe industriel de Recherche et Développement (R&D) METALTec, qui regroupe plus de 30 entreprises, et a été financé par le Bureau de Recherche et Développement Énergétique (Canada) ainsi que par le Centre Québécois de Recherche et de Développement de l'Aluminium (Québec).

J'aimerais remercier le CTA du Conseil National de Recherches Canada (CNRC) pour m'avoir offert cette opportunité, ainsi que l'ensemble de ses employés, et plus spécifiquement François Nadeau, Fatemeh Mirakhorli, Martin Larouche, Alban Morel, Siyu Tu et Marc-Olivier Gagné pour leur expertise et leur contribution, directe ou indirecte, à la réalisation de ce projet.

Je tiens à exprimer ma gratitude envers mon directeur Julien Maitre pour son encadrement, et surtout envers mon co-directeur Alexandre Gariépy, dont le suivi constant et l'accompagnement tout au long du projet ont été déterminants pour la réussite de ce travail.

Je souhaite également remercier ma famille et mes amis, plus particulièrement mes parents, pour leur soutien et leurs encouragements constants tout au long de mon parcours scolaire, m'incitant à poursuivre des études dans les domaines qui me passionnent.

Mes remerciements vont aussi à l'équipe d'athlétisme des INUK pour leur soutien financier, ainsi que pour la motivation supplémentaire qu'apporte le sport universitaire à poursuivre mes études.

CHAPITRE I

INTRODUCTION

Le soudage est un procédé industriel qui permet d'assembler plusieurs pièces métalliques pour différentes applications. Il est utilisé dans plusieurs secteurs, tels que l'automobile, l'aéronautique ou la fabrication d'équipements industriels (American Welding Society, 2025). Il existe un grand nombre de procédés de soudage, avec chacun ses propres avantages et domaines d'application. Parmi les différentes techniques de soudage disponibles, le procédé GMAW a été choisi pour ce projet en raison de sa rapidité, de sa facilité d'automatisation et de sa polyvalence avec plusieurs matériaux, comme l'aluminium (Pattanayak & Sahoo, 2021). La Figure 1.1 démontre ce à quoi ressemble une soudure GMAW.

FIGURE 1.1 : Exemple de soudure GMAW.
Image provenant de homedepot.com

Le principe du procédé GMAW repose sur la création d'un arc électrique entre les pièces à souder et un fil d'apport qui sert d'électrode. Le courant, qui est généralement élevé, permet

la fusion du fil d'apport et des matériaux de base, ce qui forme ainsi une soudure solide après le refroidissement du métal. Le courant utilisé en GMAW peut être soit continu, soit pulsé, selon les besoins du procédé et le type de matériau à souder. En mode pulsé, des pics de courant sont appliqués à intervalles réguliers, ce qui permet de contrôler plus précisément la chaleur transmise au matériau et le transfert du matériau d'apport. Cela aide à réduire les risques de déformation et permet d'avoir une meilleure maîtrise de la pénétration de la soudure, surtout sur des matériaux fins ou sensibles comme l'aluminium (Zheng *et al.*, 2022).

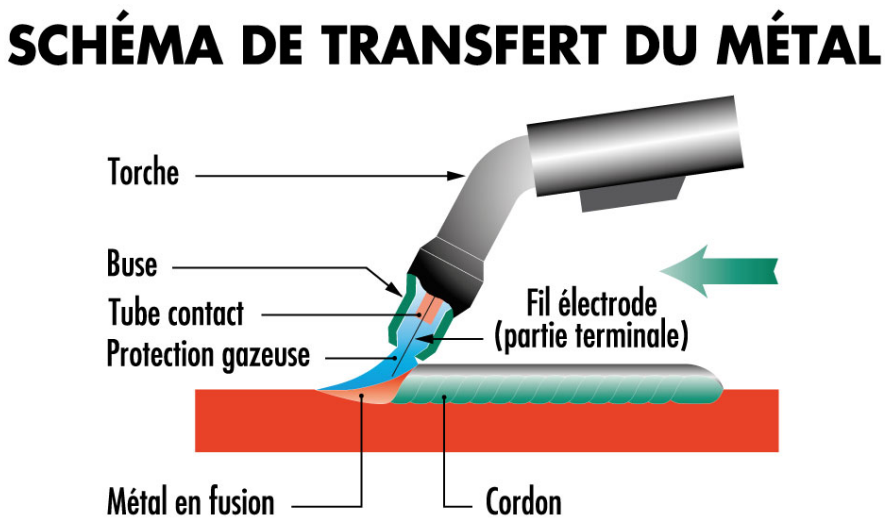


FIGURE 1.2 : Composants d'une torche de soudage GMAW.
Image provenant de easyweldfrance.com

Le fil d'apport (fil électrode ou partie terminale sur la Figure 1.2), qui est le principal conducteur du courant, est choisi selon les matériaux à assembler. Il est alimenté automatiquement par la torche de soudage qui est tenue par un opérateur ou montée sur un bras robotisé. Historiquement, le soudage GMAW était réalisé manuellement, avec un soudeur qualifié qui contrôlait visuellement et auditivement le bain de fusion. Cette opération nécessite le port d'équipements de protection individuelle, notamment un masque de soudage spécialisé (voir la Figure 1.3) qui est conçu pour protéger contre l'intense lumière émise par l'arc électrique,

ce qui inclut des rayonnements ultraviolets et infrarouges, qui sont dangereux pour les yeux et la peau.

FIGURE 1.3 : Soudeur effectuant manuellement une soudure GMAW. Image provenant de topweld.ca

Avec la pénurie de main-d'œuvre spécialisée et le besoin de répétabilité dans les productions industrielles à grand volume, il est devenu fréquent d'avoir recours à des robots automatisés pour effectuer la tâche. Ces robots se chargent d'effectuer des déplacements de la torche selon des trajectoires définies préalablement et de façon très répétable (Hadžihafizović, 2023). Cependant, il est important de différencier le robot de la source de soudage : la source contrôle les paramètres électriques (courant, tension, mode), tandis que le robot effectue uniquement les mouvements de la torche. Les deux doivent toutefois opérer ensemble pour bien réaliser les opérations.

Cette automatisation amène de nouveaux défis dans l'industrie. Contrairement au soudage manuel, il n'y a plus d'opérateur pour surveiller directement la qualité du bain de fusion

FIGURE 1.4 : Exemple de configuration avec robot (en bleu) et source (en rouge). Image provenant de lorch.eu

tout au long de la soudure. Cela fait en sorte qu'un défaut peut passer dans la chaîne de production et ne se révéler qu'à la fin lors de tests destructifs ou non destructifs, par échantillonnage (Bouslah *et al.*, 2016). Cela montre que la supervision automatique en temps réel est une capacité très importante dans des environnements robotisés.

Les récentes avancées en matériel informatique et en intelligence artificielle ont permis d'amener la fabrication intelligente pour aider à résoudre les défis mentionnés précédemment (Wuest *et al.*, 2016). Celle-ci repose sur des procédés surveillés dynamiquement, où des systèmes de détection automatisée peuvent intervenir en direct pour alerter l'opérateur, voire ajuster ou corriger les paramètres du procédé (Wu *et al.*, 2017). Cela a pour but de compenser l'absence ou une supervision minimale par un opérateur en production et à améliorer la qualité dès l'exécution du procédé.

Ce projet s'inscrit dans cette optique, en collaboration avec le CTA du CNRC, un centre de recherche appliquée qui soutient l'innovation dans l'industrie de la transformation de l'aluminium. Le CTA dispose d'installations permettant de simuler des environnements industriels réalistes dans un cadre de R&D (Fondation canadienne pour l'innovation, 2025).

1.1 PROBLÉMATIQUE

Présentement, le contrôle qualité des soudures d'aluminium repose encore souvent sur des méthodes effectuées en fin de production et qui demandent beaucoup de temps à des experts, telles que les analyses non destructives aux rayons X ou les coupes micrographiques destructives. Bien qu'elles soient fiables, ces méthodes ne permettent pas de détecter rapidement et d'intervenir directement dans la chaîne de production, ce qui réduit l'efficacité du processus et augmente les coûts. L'objectif est donc de détecter les défauts typiques de soudure en quasi-temps réel à partir des données qui peuvent être collectées pendant le procédé, telles que les signaux électriques (courant, tension) et acoustiques.

Dans le présent travail, l'explicabilité des modèles de classification est un point important pour comprendre comment les anomalies ont été détectées et faciliter l'ajustement rapide des procédés, ce qui est crucial pour améliorer la qualité et l'efficacité du soudage.

Les travaux réalisés dans le cadre de ce projet ont été effectués à l'aide d'une cellule robotisée, équipée d'une source de soudage Fronius CMT TransSynergic 4000, réputée pour sa stabilité de l'arc en mode pulsé (Viviani *et al.*, 2019), ainsi que d'un robot Yaskawa Motoman d'une capacité de charge de 45 kg, utilisé pour automatiser les déplacements de la torche le long de trajectoires linéaires simples programmées. Le cadre d'analyse serait toutefois facilement transposable à d'autres équipements.

1.2 OBJECTIFS

L'objectif général du projet est de concevoir et d'implanter un système informatique de surveillance du procédé de soudage automatisé de l'aluminium pour en démontrer les bénéfices potentiels aux participants du groupe de R&D. Une démonstration hors ligne a posteriori des capacités de surveillance de segments de soudure de l'ordre de 20 secondes est visée, plutôt qu'une surveillance en temps réel.

Les objectifs spécifiques sont les suivants :

— À la fin d'un segment de soudure, extraire les données électriques et acoustiques générées pendant le procédé de soudage à l'aide d'un système adapté ;

— Entraîner et évaluer une variété de modèles de ML pour la détection d'anomalies dans ces données ;

— Implémenter les meilleurs modèles dans un environnement représentatif de la production industrielle, en considérant les contraintes réelles telles que la latence, la robustesse et la variabilité du procédé.

Ce système a pour objectif d'offrir les fondements d'un outil de diagnostic rapide, non destructif et réactif, afin d'améliorer la qualité des soudures et de diminuer les coûts associés au contrôle qualité. Il représente la première étape vers une éventuelle extension à l'ajustement et aux corrections automatiques.

CHAPITRE II
REVUE DE LITTÉRATURE

Ce chapitre présente une revue des travaux scientifiques qui pourront servir à la mise en place du système de détection d'anomalies dans le procédé de soudage GMAW à partir de données temporelles. Cette revue est séparée en trois sections principales : l'analyse de signaux temporels (section 2.1), les approches de détection d'anomalies (section 2.2) et la combinaison des deux appliquée au contexte du soudage (section 2.3).

2.1 ANALYSE DE SIGNAUX TEMPORELS

L'analyse de signaux temporels est une étape importante pour extraire de l'information à partir de données de capteurs telles que le courant, la tension ou les émissions acoustiques. Cette section présente premièrement les méthodes d'analyse dans un contexte général, et se concentre par la suite sur leurs applications spécifiques au domaine du soudage.

2.1.1 MÉTHODES GÉNÉRALES

L'article de Christ *et al.* (2018) présente une méthode d'extraction de caractéristiques pour les séries temporelles avec le package Python TSFresh. La démarche se fait en trois étapes : l'extraction automatique de caractéristiques temporelles, fréquentielles et non linéaires, l'évaluation statistique des attributs pour déterminer lesquelles sont les plus pertinentes, ainsi qu'une sélection finale des caractéristiques qui est basée sur le contrôle du taux de fausses découvertes pour éviter d'avoir des caractéristiques qui ont été choisies par hasard. Les résultats expérimentaux obtenus sur des bases de données publiques ainsi qu'un cas industriel

(laminage d'acier) montrent une bonne capacité à extraire des caractéristiques temporelles, malgré une certaine redondance des attributs.

Dans le contexte de l'analyse de signaux temporels, TSFresh amène un grand avantage : une grande diversité de caractéristiques, ce qui peut aider à identifier des irrégularités ou des corrélations dans les signaux analysés. Toutefois, il y a certaines limites : la génération de nombreuses caractéristiques peut produire une redondance élevée, ce qui peut complexifier l'interprétation de celles-ci, ainsi que les coûts en puissance de calcul et en mémoire qui peuvent être significatifs, ce qui peut représenter une contrainte dans certains contextes industriels.

L'article de Lubba *et al.* (2019) présente une méthode d'extraction automatique de caractéristiques à partir de séries temporelles, appelée catch22, composée de 22 descripteurs sélectionnés parmi plus de 4700 caractéristiques du package hctsa. Cette sélection s'appuie sur une évaluation empirique de la performance de chaque caractéristique sur 93 tâches de classification issues de la base UEA/UCR. Le pipeline proposé comprend un filtrage statistique pour écarter les caractéristiques équivalentes à du bruit, une sélection selon la performance moyenne normalisée sur toutes les tâches, ainsi qu'une réduction de la redondance par regroupement corrélé des caractéristiques. Le résultat est un ensemble réduit, performant et rapide, qui conserve environ 90 % de la performance du jeu complet tout en étant environ 1000 fois plus rapide à calculer.

Dans le contexte de l'analyse de signaux temporels, catch22 offre plusieurs avantages pratiques : il extrait directement des caractéristiques à partir des séries temporelles brutes, sans dépendre d'une transformation préalable ni d'un grand nombre de paramètres, et les descripteurs sont conçus pour être interprétables. Cela permet d'envisager une classification rapide, robuste et compréhensible dans des contextes industriels où les contraintes de temps

réel et de calcul sont fortes. Toutefois, comme l'ensemble est fixe et non optimisé pour des cas d'usage particuliers, il peut être moins performant que des approches sur-mesure dans certains contextes spécifiques, ou pour des signaux très atypiques.

L'article de Zheng *et al.* (2014) introduit une architecture de CNN pour la classification de séries temporelles multivariées : le Multi-Channels Deep Convolutional Neural Network (MC-DCNN). Chaque dimension (canal) est traitée individuellement par un CNN 1D composé de couches de filtrage, d'activation (sigmoïde) et de pooling, puis les représentations extraites sont fusionnées et transmises à un perceptron multicouche (MLP) pour la classification finale. L'architecture est testée sur deux ensembles de données réelles : PAMAP2 (activités physiques), BIDMC (données d'électrocardiogrammes). Les résultats montrent que le MC-DCNN a obtenu 93,36 % d'exactitude sur PAMAP2 et 94,67 % sur BIDMC, ce qui est meilleur que les performances des méthodes classiques telles que 1-Nearest Neighbor avec alignement temporel dynamique (≈84 % et 92,90 % respectivement).

Dans le contexte de l'analyse de signaux temporels, cette approche est particulièrement pertinente. Elle permet de traiter séparément plusieurs sources de données, tout en apprenant une représentation conjointe pour la classification, sans extraction préalable de caractéristiques.

L'article de Fawaz *et al.* (2019) propose une revue des approches de classification de séries temporelles par DL. Les auteurs comparent une dizaine d'architectures de bout en bout, ce qui inclut des réseaux de type perceptron multicouche, CNN, ResNet, MC-DCNN, ainsi que des réseaux récurrents (Long Short-Term Memory (LSTM), unité récurrente fermée) et réservoirs (Echo State Networks), entraînés sur 97 jeux de données issus de la base UEA/UCR et d'autres sources. Tous les modèles sont évalués sans extraction manuelle de caractéristiques. Le protocole expérimental est le suivant : chaque modèle est entraîné 10 fois et les performances sont comparées avec des tests statistiques (Friedman et Wilcoxon-Holm) pour

garantir la robustesse des résultats. ResNet se démarque comme l'architecture profonde la plus performante, avec une exactitude statistiquement équivalente à celle du classificateur HIVE-COTE qui ne fait pas partie du DL, mais avec une complexité computationnelle très inférieure.

Cette revue permet d'appuyer l'utilisation de modèles de DL comme ResNet directement sur des signaux temporels non prétraités. Elle met en avant leur capacité à apprendre automatiquement des représentations discriminantes sans ingénierie manuelle. Toutefois, ces architectures demandent généralement beaucoup de données pour un entraînement optimal et demandent plus de ressources pour l'entraînement que des méthodes de ML classique.

2.1.2 APPLICATIONS AU PROCÉDÉ DE SOUDAGE

L'article de Liang *et al.* (2021) étudie les caractéristiques des signaux acoustiques générés lors du procédé de soudage automatique pulsé GMAW. Un prétraitement sur des données acoustiques est effectué, comprenant la suppression de la composante continue, un fenêtrage temporel (10-30 ms) et une réduction du bruit par ondelettes. L'analyse du signal repose sur deux méthodes : la densité spectrale de puissance calculée via la méthode de Welch, et une décomposition en paquets d'ondelettes. Les résultats montrent que la distribution fréquentielle de l'énergie acoustique varie selon l'état de pénétration du bain de fusion : sous-pénétration, pénétration complète ou sur-pénétration. Plus précisément, lorsque l'apport de chaleur augmente, l'énergie se déplace des hautes fréquences (7-10 kHz) vers les bandes plus basses (2-4 kHz).

Ces observations indiquent que les signaux acoustiques des soudures contiendraient des informations pertinentes pour la surveillance de la qualité du soudage. Cependant, le nombre de soudures utilisées n'est pas mentionné, ce qui laisse supposer qu'il pourrait être plutôt bas.

L'article de Thekkuden *et al.* (2020) présente une analyse statistique de la tension et du courant enregistrés lors du soudage GMAW, en fonction de trois paramètres du procédé : la longueur d'arc, le débit de gaz de protection et la vitesse d'avance. Les signaux électriques sont analysés globalement (sans segmentation) pour extraire quatre caractéristiques : la moyenne et l'écart-type de la tension et du courant. Le mode de soudage n'est pas mentionné, mais les résultats montrent que la tension est principalement influencée par la longueur d'arc et le débit de gaz de protection, tandis que le courant est sensible à la longueur d'arc et à la vitesse d'avance. Des modèles de régression quadratique sont proposés afin de modéliser la relation entre les paramètres de soudage et les caractéristiques statistiques des signaux électriques.

Dans le cadre d'analyse de données temporelles, cette étude montre l'intérêt de la moyenne et de l'écart-type de la tension et du courant comme indicateurs de stabilité de l'arc.

2.2 DÉTECTION D'ANOMALIES

La détection d'anomalies est une composante essentielle pour concevoir un système de surveillance industrielle. Cette section commence par montrer des techniques utilisées dans le secteur manufacturier en général, avant de se concentrer sur le procédé de soudage.

2.2.1 DANS LE SECTEUR MANUFACTURIER

L'article de Scholz *et al.* (2024), extrait du livre Artificial Intelligence in Manufacturing, montre des méthodes de détection d'anomalies dans le secteur manufacturier. Après une présentation des concepts fondamentaux, les auteurs détaillent les différentes techniques :

— **Méthodes statistiques** : clustering (k-means, espérance-maximisation), détection de valeurs aberrantes (méthodes basées sur les plus proches voisins), classification supervisée

(arbres de décision, logique floue, classification naïve bayésienne, machine à vecteurs de support, algorithmes génétiques).

— **DL** : auto-encodeurs et en particulier ceux de type LSTM pour les séries temporelles multivariées.

Les auteurs mentionnent que l'efficacité de la détection repose notamment sur le choix des variables pertinentes provenant des capteurs (pression, température, énergie), ainsi que sur des techniques de prétraitement comme le débruitage et l'analyse fréquentielle.

Une étude de cas présente une application dans le cadre du projet européen knowlEdge, sur un processus de moulage par soufflage de réservoirs à carburant. Les données venant de plus de 100 capteurs sont segmentées par cycles machine. Un auto-encodeur LSTM est entraîné pour encoder chaque cycle, et les anomalies sont détectées par l'erreur de reconstruction, comparée à un seuil basé sur une distribution normale.

Cet article met bien en évidence la diversité des approches existantes, et montre l'intérêt des auto-encodeurs LSTM pour la détection d'anomalies non supervisée sur des séries temporelles industrielles. Toutefois, les résultats restent qualitatifs, sans comparaison quantitative sur d'autres procédés comme le soudage.

L'article de Elía & Pagola (2025) propose une revue des techniques de détection d'anomalies appliquées à l'industrie manufacturière. Il présente un cadre de classification des scénarios industriels, puis évalue 16 algorithmes sur 29 jeux de données réelles couvrant des domaines variés (moteurs, robots, etc.). Les algorithmes testés incluent des approches de ML classique (forêt d'isolation, Principal Component Analysis (PCA), facteur local d'anomalie, k-plus proches voisins) et de DL (CNN, LSTM, DeepSVDD, DevNet). Des méthodes non-supervisées (auto-encodeurs, auto-encodeurs variationnels), semi-supervisées (XGBOD, GANomaly) et supervisées (RF, CatBoost) sont aussi incluses dans ces algorithmes.

L'étude montre que les méthodes par ensemble, comme XGBOD (semi-supervisée), CatBoost et RF (supervisées), obtiennent les meilleures performances sur les grandes catégories industrielles testées : les moteurs et transmissions, les robots et véhicules autonomes, l'usinage et la fabrication additive, ainsi que les processus multi-étapes intégrés. Cependant, les méthodes profondes non-supervisées (DeepSVDD, LSTM Outlier Detector) ont eu des résultats plus variables. Les auto-encodeurs et les réseaux LSTM restent toutefois adaptés à la détection d'anomalies dans les séries temporelles longues ou complexes.

Cet article illustre l'importance grandissante de l'IA dans l'industrie de la fabrication et constitue une très bonne référence pour explorer des algorithmes de détection d'anomalies en fonction du type de capteur utilisé et de la nature des anomalies. Sa principale limite pour le contexte actuel est l'absence de procédé de soudage, à l'exception du Wire Arc Additive Manufacturing (WAAM) pour lequel les réseaux de neurones semblaient plus couramment étudiés.

2.2.2 APPLICATIONS AU PROCÉDÉ DE SOUDAGE

L'article de Huang *et al.* (2023) propose une méthode de détection de défauts de soudure à partir d'images aux rayons X, basée sur un réseau de CNN multi-échelles amélioré (IMSACNN) qui contient un Convolutional Block Attention Module (CBAM). L'architecture utilise des modules Inception pour capter des caractéristiques à plusieurs échelles, tandis que le CBAM affine la sélection des régions pertinentes.

Cette architecture est testée sur un ensemble de 3000 images haute résolution, dont 2000 comportaient quatre types de défauts (fissures, porosités, manque de fusion, manque de pénétration). Les résultats obtenus atteignent une exactitude globale de 87,19 %, avec des taux

de reconnaissance individuels allant de 94 % à 100 %. Le modèle est comparé à des modèles de référence (CNN, CNN avec apprentissage par transfert, DL), qu'il surpasse en exactitude.

Cela montre qu'un CNN amélioré peut être utilisé pour classifier des défauts de soudure dans des images aux rayons X, et ce, avec 4 types de défauts en même temps. Cependant, aucune donnée temporelle venant de capteur n'est utilisée pour chercher et identifier des défauts à la source.

L'article de Hong *et al.* (2023) présente une méthode de surveillance en temps réel de la qualité des soudures de tôles ultra-fines en acier inoxydable 304 (épaisseur $\leq$ 0,15 mm) réalisées par soudage à l'arc pulsé plasma micro (MPAW). Le système a un dispositif de microvision qui capture en temps réel la zone du bain de fusion et permet d'extraire des caractéristiques, telles que la largeur, le périmètre et la surface.

Ces caractéristiques servent à entraîner un modèle de machine à vecteurs de support qui classifie les soudures en trois états : manque de fusion, formation de bosses (humping), et soudure acceptable. Les résultats montrent une exactitude de 96,09 % sur les données de test et de 94,98 % sur des données non vues lors de l'entraînement.

Bien que cette étude ne soit pas du soudage GMAW et n'utilise pas de données temporelles, elle démontre tout de même une classification de deux anomalies différentes en temps réel grâce au modèle de machine à vecteurs de support.

Le cinquième article inclus dans la thèse de doctorat d'Ahmad Aminzadeh (Aminzadeh, 2024) propose une méthode de détection en temps réel de la porosité dans le soudage laser de l'aluminium, en se basant sur l'analyse de la morphologie 3D du trou de serrure (cavité générée par le laser), observée par une caméra haute vitesse. Chacune des images capturées est segmentée afin d'extraire automatiquement des caractéristiques du trou de serrure (aire,

symétrie, orientation, etc.) à l'aide du logiciel ImageJ. L'apprentissage est réalisé à partir d'un modèle de RF et les caractéristiques les plus importantes sont analysées après les tests.

Les résultats montrent un F1-Score de 83 % pour distinguer les soudures avec ou sans porosité. Bien que l'étude a obtenu des résultats positifs, il s'agit de classification binaire pour le soudage au laser qui est différent du GMAW. De plus, les méthodes utilisant la vision nécessitent une ligne de vue sur le bain de fusion, ce qui peut être un défi dans certaines applications industrielles.

2.3 APPROCHES INTÉGRÉES D'ANALYSE DE SIGNAUX ET DE DÉTECTION D'ANOMALIES EN SOUDAGE

Cette section regroupe des études où l'analyse de signaux temporels est utilisée pour la détection des anomalies dans le procédé de soudage. Elle met de l'avant les types de données temporelles utilisées, les techniques de prétraitement utilisées, les méthodes de détection d'anomalies ainsi que les différentes anomalies détectées.

L'article de Mattera & Nele (2025) propose une méthode de détection d'anomalies en temps réel dans un contexte de WAAM, appliquée à l'alliage Invar 36. L'objectif est de comparer plusieurs approches d'apprentissage supervisées, semi-supervisées et non supervisées pour détecter automatiquement des anomalies de procédé à partir de signaux électriques. Les données de tension et de courant sont acquises à 5 kHz, puis découpées en fenêtres de 1 s. Un total de 83 caractéristiques est extrait dans les domaines temporel, fréquentiel (Fast Fourier Transform (FFT)) et temps-fréquence (transformée de Morlet et ondelettes de Daubechies). Le jeu de données contient 1698 échantillons, dont 344 (20 %) sont étiquetés comme anormaux. Les anomalies regroupent plusieurs défauts : porosité, affaissement de couche, projections

excessives, instabilités d'amorçage ou décalage du bain de fusion, mais ils sont tous regroupés sous une même classe, ce qui représente donc un problème de classification binaire.

Le réseau de neurones a obtenu le meilleur F1-Score (92,1 %), tandis que la forêt d'isolation a obtenu le meilleur F2-score (91,5 %). Le F1-Score mesure l'équilibre entre précision et rappel, tandis que le F2-score donne davantage de poids au rappel, ce qui le rend plus adapté lorsque l'on cherche à réduire le risque de manquer une anomalie, même si cela entraîne plus de faux positifs. Ces résultats sont très prometteurs, mais il est important de prendre en compte que la classification s'est faite avec seulement deux classes : le modèle ne fournit donc pas d'indication sur le type d'anomalie rencontrée. De plus, cet article parle de WAAM, qui est une utilisation du GMAW pour faire de l'impression 3D et non des soudures de deux pièces métalliques.

L'article de Jin *et al.* (2020) présente un modèle de CNN pour détecter un défaut interne difficilement observable dans les soudures GMAW en mode de transfert en court-circuit à tension constante : une mauvaise formation du cordon arrière (back-bead). Des signaux de courant sont mesurés à haute fréquence (50 kHz) durant le procédé, puis transformés en scalogrammes, c'est-à-dire en représentations temps-fréquence sous forme d'images 2D, obtenues par transformée en ondelettes. Ces images sont utilisées comme entrée d'un CNN composé de seulement deux couches de convolution. Au total, l'ensemble de données contient 1 667 soudures de 128x128 pixels avec trois canaux de couleur (RGB).

Le modèle atteint en moyenne une exactitude de 93,5 % et démontre qu'il est possible de détecter des anomalies seulement avec un signal de courant électrique. Toutefois, la classification est binaire, ce qui fait que la présence d'autres anomalies différentes pourrait influencer les résultats.

L'article de Cullen & Ji (2025) propose un système de détection de défauts dans le procédé GMAW, basé uniquement sur l'analyse du signal acoustique capté par un microphone directionnel positionné à 50 cm de la zone de soudage. Le signal, échantillonné à 96 kHz, est traité dans le domaine fréquentiel pour identifier le mode de transfert (court-circuit, globulaire, pulvérisation). Les anomalies sont ensuite détectées à partir de déviations anormales de ces modes de transfert. Cette approche ne se base pas sur un algorithme d'Intelligence Artificielle (IA), mais sur des règles heuristiques empiriques qui sont faites à partir de signatures acoustiques typiques de chaque défaut.

Plutôt qu'un classifieur supervisé multiclasse, le système applique des règles distinctes pour chaque type de défaut, en les comparant uniquement à des soudures normales (équivalent à de la classification binaire). Les brûlures (burn-through) sont détectées à partir de perturbations prolongées du spectre acoustique, avec une détection correcte dans 90 % des cas. La porosité est identifiée grâce à des pics acoustiques transitoires spécifiques, détectés à 100 % dans les tests. La pénétration n'est pas détectée comme un défaut, mais plutôt estimée en continu par régression linéaire sur la fréquence dominante du signal, avec une erreur inférieure à 15 %.

Cette approche montre que de simples caractéristiques de données acoustiques peuvent être suffisantes pour détecter des anomalies, et ce, même sans modèle de ML. Par contre, les anomalies ont été détectées de manière "binaire", ce qui indique que les résultats auraient pu être moins bons avec des ensembles de données plus complexes. De plus, la nature empirique et spécifique des règles fait en sorte qu'elles ne seraient peut-être pas applicables dans un autre environnement de soudage.

L'article de Duongthipthewa *et al.* (2023) présente un système de surveillance d'un procédé robotisé de soudage industriel qui se base sur des données de courant et de tension électrique. Ces signaux sont transformés en spectrogrammes de 432x288 pixels, qui servent

à entraîner un CNN. Le système doit classifier l'état de la soudure selon trois classes qui représentent des longueurs d'arc (15 mm, 10 mm, 20 mm) et qui reflètent l'apparence et la pénétration, la classe de 15 mm étant la référence.

Les résultats indiquent une exactitude de 89 %, avec un F1-Score allant jusqu'à 100 % pour les bonnes soudures (15 mm). Toutefois, la petite taille de l'ensemble de données (180 échantillons) et le nombre de classes très bas amène une certaine limite à une application en contexte réel.

L'étude de Skar *et al.* (2023) présente un modèle non supervisé d'auto-encodeur LSTM pour la détection d'anomalies dans les signaux acoustiques enregistrés lors du procédé Flux-Cored Arc Welding (FCAW), semblable au GMAW. Le modèle est entraîné à reconstruire des signaux acoustiques correspondant à des soudures normales. Les anomalies sont ensuite identifiées lorsqu'un écart de reconstruction dépasse un certain seuil. Les signaux sont enregistrés à 192 kHz, sans traitement complexe (aucun filtrage fréquentiel, utilisation du signal brut seulement). Les défauts incluent l'absence de gaz de protection, une mauvaise compensation de la longueur d'arc, et des variations de trajectoire (weaving). Ces défauts sont détectés à travers une augmentation de l'erreur de reconstruction comparé aux données normales.

L'approche permet bien de détecter les anomalies, mais ne permet pas de les classifier (classification binaire). De plus, une seule soudure par type d'anomalie et une soudure de référence ont été utilisés pour les tests, ce qui est très bas pour dire qu'un modèle est efficace.

L'article de Schmidt *et al.* (2021) propose un système d'observation en temps réel du désalignement (offset) de l'outil de soudage robotisé. 5 types de données temporelles du procédé sont utilisés : courant, tension, vibrations mécaniques, émissions acoustiques et lumière d'arc. Les signaux sont acquis à haute fréquence (200 kHz) et découpés en segments de 100 ms. Des caractéristiques temporelles sont extraites avec la librairie TSFresh, tandis

qu'une FFT suivie d'une réduction de dimension par PCA est appliquée aux composantes fréquentielles. Chaque segment est ainsi représenté par un vecteur de 131 attributs par capteur. Trois modèles de régression sont entraînés à partir d'un ensemble de 1215 soudures et comparés : régression linéaire, régression Elastic Net et réseau de neurones. Le réseau de neurones a obtenu les meilleures performances, avec une erreur moyenne absolue de 0,053 mm (écart-type de 0,044 mm).

Cette approche permet de bien prédire le désalignement du robot de soudage et elle démontre l'utilisation de plusieurs capteurs en simultané avec plusieurs méthodes intéressantes d'extraction de caractéristiques. Cependant, bien qu'il y ait eu des tests avec un seul capteur à la fois, il n'y a pas de comparaison de performances entre ceux-ci.

L'article de Rohe *et al.* (2021) présente une méthode de détection d'anomalies à l'aide d'un CNN, à partir de l'analyse de signaux acoustiques mesurés pendant le processus de soudage GMAW. Les signaux acoustiques sont enregistrés, avec une fréquence d'échantillonnage de 48 kHz, lors de soudures avec différents débits de gaz de protection (0, 50, 60, 70, 80, 90 et 100 %), dans le but prédire si le débit est trop bas ou normal. Un prétraitement est appliqué sur les signaux acoustiques, ce qui inclut une segmentation temporelle (2 s par segment), une transformée de Fourier à court terme, suivie de l'application d'une banque de filtres Mel. Les spectrogrammes obtenus sont ensuite utilisés comme entrées pour classifier avec le CNN.

Les résultats montrent que le CNN atteint une exactitude moyenne de 84 %, ainsi qu'une exactitude de 100 % pour le débit de gaz à 0 % et une exactitude de 94 % pour le débit de gaz à 100 %. Plusieurs autres modèles ont été entraînés en retirant une bande de fréquence à chaque fois lors du prétraitement (leave-one-out) et cela a démontré que 3 bandes de fréquences étaient plus importantes que les autres pour avoir de bonnes prédictions. Bien que cette étude ne se concentre que sur un type d'anomalie, elle fait tout de même de la classification multiclasse

pour les différents débits de gaz. Il faut aussi mentionner que le nombre de segments utilisés est de 302, donc le nombre réel de soudures est bien inférieur à cela, ce qui est plutôt bas pour un modèle de DL.

2.4 CONCLUSION

La revue de littérature présentée dans ce chapitre a permis d'établir l'état de l'art pour la mise en place d'un système intelligent de détection d'anomalies appliqué au soudage.

Premièrement, l'analyse de signaux temporels est une approche démontrée pour extraire des informations utiles à partir de données issues de capteurs industriels. Des outils tels que TSFresh ou catch22 permettent d'automatiser cette étape, et amènent des compromis entre diversité des caractéristiques, rapidité de calcul et interprétabilité. De plus, les méthodes de DL comme les CNN se sont révélées particulièrement efficaces pour traiter directement des signaux bruts à condition de disposer de volumes de données suffisants.

La détection d'anomalies a beaucoup de techniques différentes, en commençant par des approches statistiques classiques à des modèles profonds non supervisés, comme des auto-encodeurs. Ces derniers sont particulièrement adaptés aux séries temporelles complexes. Les méthodes supervisées par ensemble, telles que les forêts aléatoires ou CatBoost, ont aussi montré des performances robustes dans plusieurs cas industriels.

Plusieurs travaux ont démontré l'intérêt d'utiliser conjointement des signaux temporels et des techniques de détection d'anomalies pour évaluer la qualité des soudures, en particulier en GMAW. Les données électriques et acoustiques sont parmi les plus fréquemment exploitées. Toutefois, la majorité des approches testées restent limitées à des classifications binaires (ou au mieux 3-4 classes) ou à des cas d'usage spécifiques à partir de données contrôlées, ce qui rend leur généralisation plus compliquée pour cet environnement industriel complexe.

Bien que la littérature montre plusieurs méthodes pour développer un système de détection d'anomalies pour le soudage GMAW, elle présente encore plusieurs limites. Il n'existe pas de consensus clair sur le type d'algorithme ou le type de données à utiliser. Peu d'études comparent les performances des algorithmes de ML classique à ceux du DL pour le soudage. De plus, les approches qui combinent des signaux acoustiques et électriques sont plutôt rares et aucune comparaison n'est faite sur leurs performances individuelles. Les ensembles de données publics sont aussi très rares dans ce domaine, ce qui empêche de faire des comparaisons reproductibles d'un article à l'autre. Aussi, la majorité des travaux font de la classification binaire, ou au mieux à quelques classes, et ne cherchent pas à distinguer différents types d'anomalies qui peuvent être rencontrés en conditions réelles.

Dans ce contexte, le projet présenté dans ce mémoire cherche à combler plusieurs de ces lacunes. Il propose une comparaison entre les signaux électriques et acoustiques, afin d'évaluer leur efficacité respective pour la détection d'anomalies en soudage GMAW. Le projet compare aussi les performances d'un modèle de ML classique à celles d'un modèle de DL. De plus, l'ensemble de données associé contient plusieurs types d'anomalies différents, ce qui permet de dépasser la simple classification binaire et d'évaluer la capacité des modèles à distinguer différentes anomalies en vue de guider les ajustements de procédé.

CHAPITRE III
DESCRIPTION DU PROBLÈME DE SOUDAGE

En amont, les experts de soudage du CNRC ont identifié avec leurs partenaires industriels un ensemble de 7 conditions anormales de soudage typiques à répliquer en laboratoire. La Figure 3.1 illustre schématiquement ces irrégularités qui peuvent se produire dans l'industrie. Donc, l'ensemble de données est composé de soudures dont certaines contiennent une anomalie qui a volontairement été faite pour simuler des situations anormales qui peuvent avoir lieu en industrie. Les données utilisées ont été générées dans un environnement expérimental contrôlé, plutôt que directement en production, afin de maîtriser les paramètres du procédé et d'être sûr que les anomalies sont bien étiquetées. Un plan d'expérience avait été établi, incluant des répétitions, pour un total de 60 soudures. Pour certaines anomalies, différents niveaux de sévérité de déviation du cas de référence ont de plus été testés. Bien que le nombre de 60 soudures soit considéré comme bas du point de vue ML, cela représente un effort expérimental substantiel à collecter, car chaque configuration sert à faire des soudures d'un seul type et sévérité d'anomalie spécifique. Ce processus peut être long et doit être répété pour chaque nouveau cas testé, ce qui augmente le travail nécessaire. Il est bien connu dans le monde du ML qu'un nombre de données plus grand sera généralement bénéfique, donc le nombre de soudures étant assez bas pourrait nuire à la tâche de classification, ce qui est une contrainte plutôt fréquente dans la littérature (Rohe *et al.*, 2021; Duongthipthewa *et al.*, 2023; Skar *et al.*, 2023).

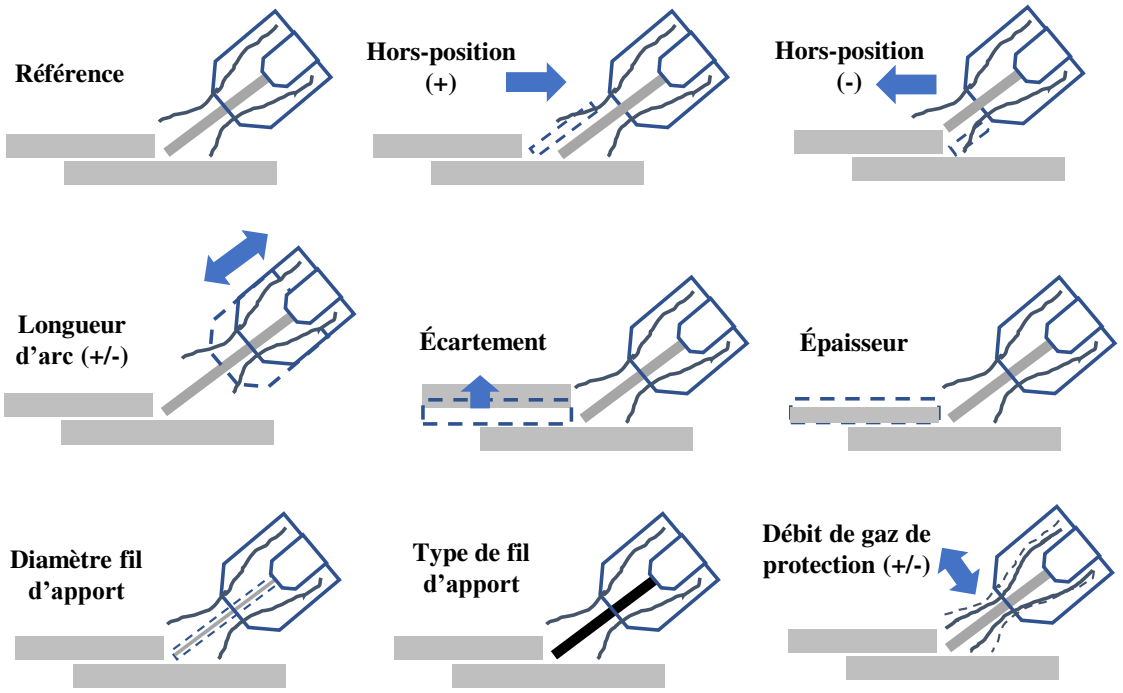


FIGURE 3.1 : Représentation visuelle des anomalies de soudures GMAW à l'étude. Adapté de Briand *et al.* (2024)

L'objectif global de notre projet est de classifier des soudures. En fonction de l'objectif pratique visé, il existe plusieurs manières de catégoriser notre ensemble de données. En effet, tel que mentionné précédemment, certaines soudures ont été volontairement réalisées avec des anomalies courantes du domaine. Cela signifie que le type de déviation est connu pour chaque spécimen, ce qui est essentiel dans notre problème numérique. Suite à cela, des évaluations par métallographie optique et par rayons X ont été faites pour vérifier si chaque soudure est conforme aux standards de l'industrie. Le test de rayons X sert à créer une image permettant d'observer la structure interne d'une soudure et ainsi d'être en mesure de détecter les défauts, tels que des porosités, des fissures, des inclusions, des manques de fusion ou des caniveaux (Wang *et al.*, 2025). Le test par métallographie optique consiste à prélever un échantillon

d'une soudure et de l'évaluer en observant au microscope une section prise aléatoirement pour détecter des défauts internes, tels que des fissures microscopiques, des anomalies dans la structure de grains ou des porosités et inclusions microscopiques (Vander Voort, 2011). Avec toutes ces informations pour chacune des soudures, il y a trois différentes méthodes de classification qui peuvent être utilisées :

Classification basée sur la procédure : Cela consiste à séparer les soudures en fonction de si elles sont des références ou des anomalies, seulement en fonction des paramètres nominaux. Les soudures sont donc classées comme étant l'un des types d'anomalies ou de type référence, seulement en prenant en compte si la soudure a été faite selon la procédure de référence. Il faut donc comprendre que les classes sont assignées sans prendre en compte la sévérité de la déviation à la procédure et si les soudures passent ou non les tests de conformité de l'industrie.

Classification basée sur les résultats des tests : Cette méthode requiert d'avoir effectué les évaluations par métallographie optique et par rayons X. Suite à cela, deux classes possibles sont assignées aux soudures : réussite ou échec. Cela permet donc de réduire la complexité du problème en le réduisant à deux classes, mais sans pouvoir connaître la cause exacte d'un échec qui permettrait de réaligner rapidement le procédé.

Classification combinée avec diagnostic : Cette méthode est une combinaison des deux autres méthodes. Elle utilise les évaluations par métallographie optique et par rayons X, et lorsqu'il y a un échec, le type d'anomalie est assigné comme classe. Il y a donc la classe de réussite ainsi que les anomalies, ce qui permet de savoir si une soudure a échoué ou non et pour quelle raison elle a échoué.

Dans ce projet, lorsqu'il y a une évaluation de qualité, le classement se fait pour la soudure entière, même lorsqu'un défaut est localisé dans une petite portion de celle-ci. Cela fait donc en sorte que des parties de soudures qui passent l'évaluation sont tout de même classifiées

comme des échecs. La modélisation en est impactée puisque le modèle d'IA pourrait être biaisé en raison d'étiquetages ne représentant donc pas la réalité, ce qui pourrait réduire sa capacité de détection des anomalies. Même s'il aurait été préférable de considérer uniquement les segments de soudures problématiques comme des échecs, cela aurait demandé un effort considérable de caractérisation traçable spatialement qui était en dehors de la portée du projet. De plus, parmi les soudures de références, l'une d'entre elles a échoué les tests de qualité, ce qui fait qu'elle a été retirée de l'ensemble de données.

Il est important de mentionner qu'il y a un seuil de tolérance, qui dépend de l'application, pour déterminer si une soudure est considérée comme une réussite ou un échec en fonction des défauts observés. Cela fait en sorte que deux soudures qui sont proches de ce seuil, une qui passe et une qui échoue les tests, pourraient être très semblables au niveau de leur signature électrique ou acoustique. Il est donc évident qu'il y a un certain niveau de confusion dans les données, ce qui augmente le niveau de difficulté pour l'entraînement des modèles.

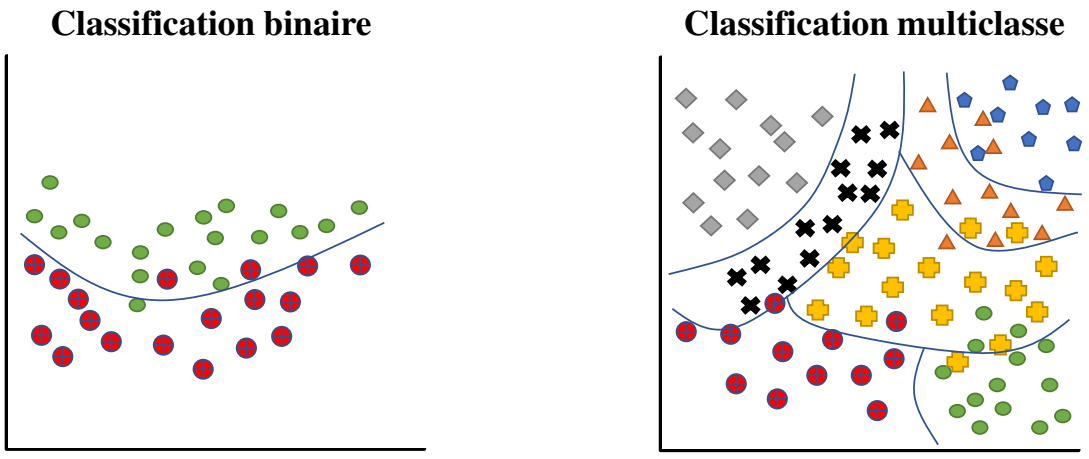


FIGURE 3.2 : Représentation schématique de la différence de complexité entre la classification binaire et multiclasse.
Adapté de Briand *et al.* (2024)

Tel qu'illustré dans la Figure 3.2, la classification multiclasse est beaucoup plus complexe que la classification binaire. En effet, la classification binaire, qui implique la séparation entre seulement deux classes distinctes, donne généralement des prédictions plus précises en raison de la simplicité de ses frontières de décision. En revanche, la classification multiclasse nécessite la résolution de plusieurs frontières complexes entre les différentes catégories. Avec environ 10 classes dans notre situation, la tâche devient donc beaucoup plus difficile pour les algorithmes en raison des différentes classes qui peuvent avoir des caractéristiques similaires. Dans une situation comme celle-ci, il faut généralement une plus grande variété et un plus grand nombre de données pour aider les modèles à différencier les classes. En prenant ces facteurs en considération, il est facile de remarquer que la situation représente un bon défi numérique, principalement en raison du volume de données relativement bas et du nombre élevé de classes, ce qui complique la tâche des algorithmes pour obtenir des performances optimales.

En plus de l'ensemble de données principal basé sur des soudures en recouvrement d'aluminium AA6061-T6 en feuilles avec un fil d'apport ER5356, un deuxième ensemble a été constitué à partir de soudures d'acier laminé galvanisé en feuilles avec un fil d'apport de grade 70S6, afin de tester la transférabilité des modèles avec des données différentes. Ces données d'acier ont été collectées de la même manière et avec les mêmes équipements et le même mode de soudage que pour l'aluminium, ce qui fait que les données ont la même forme et pourront donc être utilisées pour entraîner des modèles sans y faire de modifications.

CHAPITRE IV
MÉTHODOLOGIE ET DÉVELOPPEMENT

Ce chapitre présente la démarche choisie, en commençant par la collecte et le prétraitement des données. Suite à des tests préliminaires, le RF et le CNN ont été choisis comme algorithmes pour mettre en place les modèles de classification. Ces choix reposent sur leur performance qui a été supérieure par rapport aux autres algorithmes testés qui sont le Decision Tree, la Logistic Regression, le Linear Discriminant Analysis et le K-Nearest Neighbor en ML classique, ainsi que le Recurrent Neural Network (RNN), le LSTM et le réseau dense en DL. Ces tests n'étaient que préliminaires et ont évalué les performances de classification en termes d'exactitude (accuracy) avec les classes basées sur la procédure, mais les résultats exacts n'ont pas été sauvegardés et ne sont pas présentés ici. Cela a donc amené au choix du RF et du CNN pour continuer la suite des travaux, car ils semblaient être les plus prometteurs. Par la suite, ces deux algorithmes ont été testés selon plusieurs variations de paramètres pour évaluer leur performance sous différentes conditions et pour déterminer les paramètres optimaux.

4.1 CHOIX, COLLECTE ET PRÉTRAITEMENT DES DONNÉES

Les données utilisées dans cette étude proviennent de deux sources principales :

- Comportement électrique : Dans le but d'analyser le comportement de l'arc de soudage, un oscilloscope numérique est utilisé pour mesurer la tension et le courant à très haute fréquence (40 kHz). Ces mesures à haute fréquence sont particulièrement utiles pour les procédés pulsés utilisant un contrôle en boucle fermée, car elles permettent d'observer des pulsions avec un niveau de détail très élevé, ce qui facilite l'analyse de ces signaux.

- Comportement acoustique : Pour analyser le comportement acoustique, un microphone externe est utilisé pour enregistrer le son émis par la soudure avec une fréquence de 48 kHz. En effet, il est généralement reconnu que les soudeurs experts peuvent reconnaître la qualité à partir du bruit émis par l'arc électrique (Cullen & Ji, 2025). Les soudeurs sont capables de reconnaître au bruit certains changements, tels que les variations de tonalité et de volume, qui dépendent du mode de transfert de l'arc, de la vitesse de soudage, de l'alimentation du fil, et de la stabilité de l'arc.

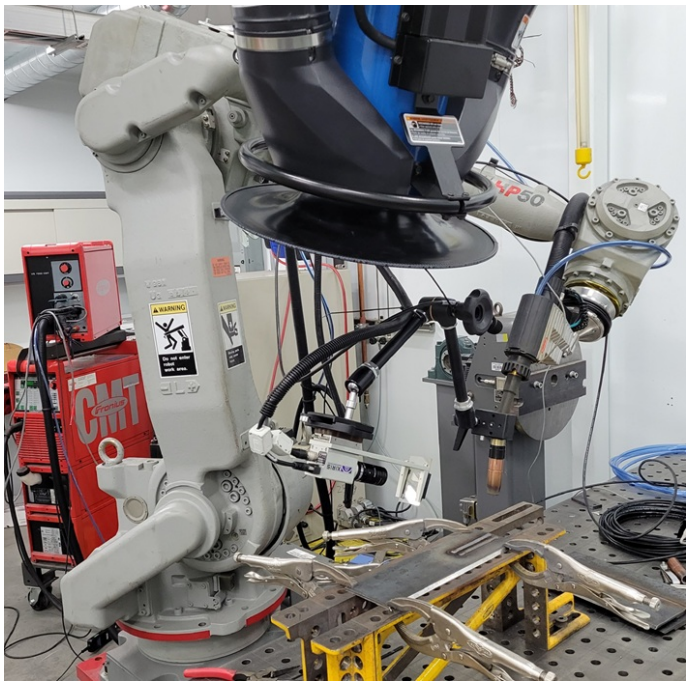


FIGURE 4.1 : Installation de soudage GMAW robotisé utilisée au CTA pour l'acquisition des données.

Les données sont donc de nature acoustique et électrique, deux types différents. De plus, un robot (voir Figure 4.1) est utilisé pour positionner et déplacer la torche de soudage et des données sont reliées à la position de ce robot en fonction du temps pour chaque soudure. Ces données ne sont toutefois pas utilisées dans ce projet en raison du fait que nous ne cherchions

pas à localiser la position d'une anomalie dans nos soudures (la soudure est une seule classe dans son entièreté). Il est aussi important de prendre en considération que certains filtres ont été appliqués en amont de la sauvegarde des données. Les filtres sont une coupure passe-bas à 4 kHz pour les données électriques et un filtre entre 1 Hz et 20 kHz pour les données acoustiques pour ainsi réduire les bruits parasites. Bien que ce filtrage permette de réduire le bruit ou les interférences dans nos données, il est important de reconnaître qu'il existe un risque potentiel de perte d'informations pertinentes pour nos modèles de classification. Par exemple, certaines anomalies subtiles dans le signal électrique ou acoustique, qui pourraient être des indicateurs d'une défaillance, pourraient se situer en dehors des plages de fréquence conservées par les filtres. Cependant, les seuils des filtres appliqués ont été déterminés par des experts-soudeurs, basés sur leur expérience et leur compréhension approfondie du procédé de soudage, ce qui vise à réduire les données parasites tout en préservant les informations essentielles à l'analyse.

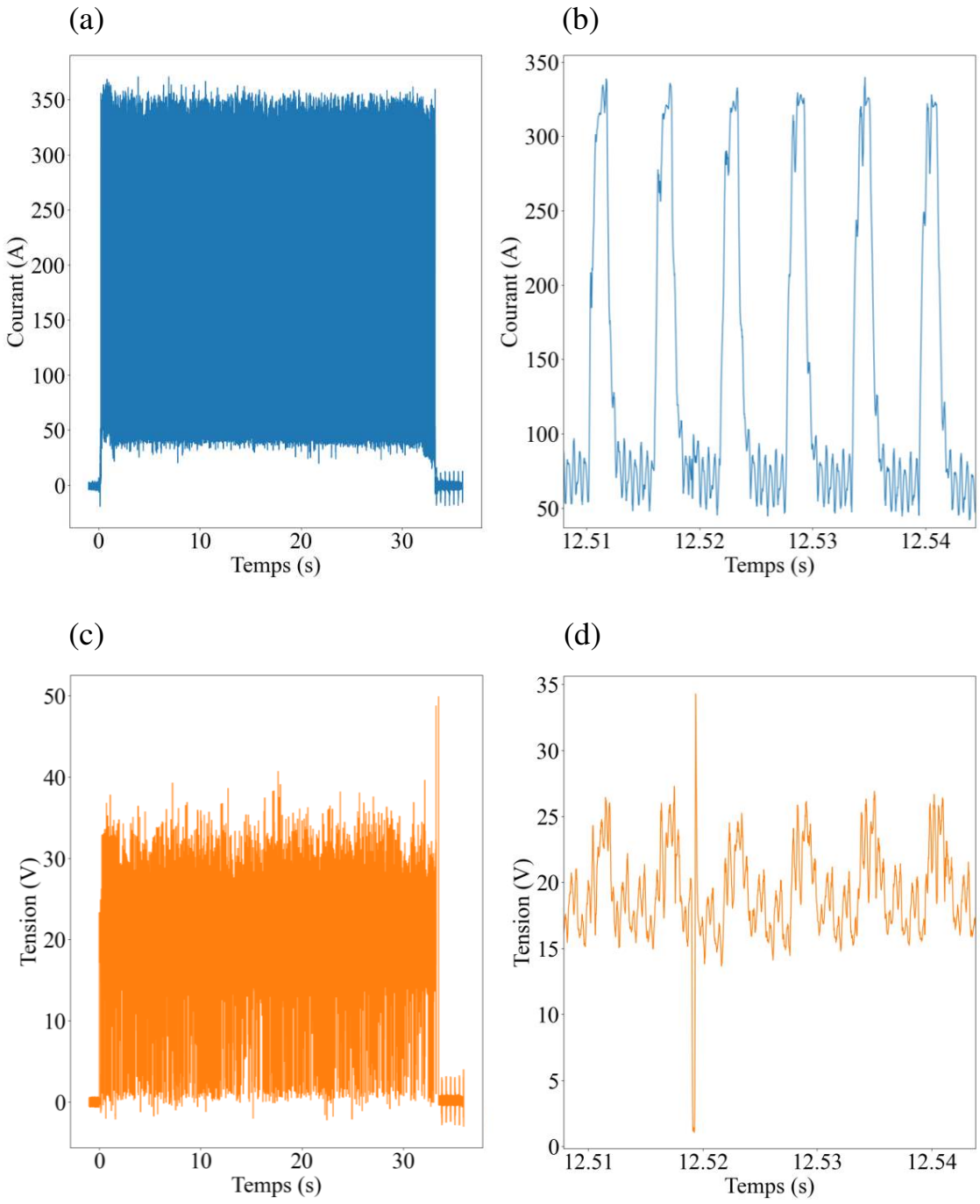


FIGURE 4.2 : Exemple de données électriques pour une soudure GMAW.

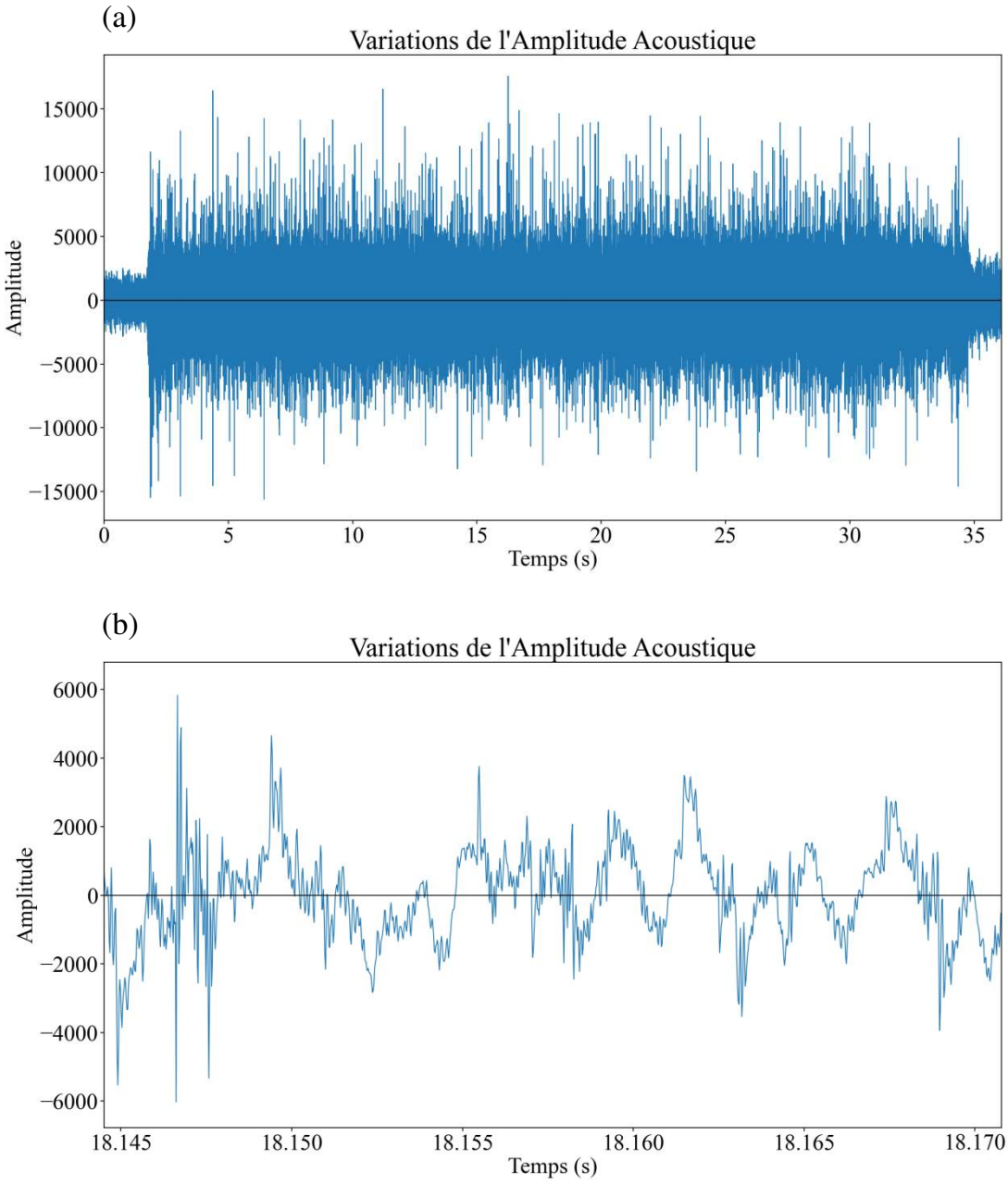


FIGURE 4.3 : Exemple de données acoustiques pour une soudure GMAW.

Lorsque les données électriques sont observées à une échelle de quelques centièmes de seconde, tel que démontré dans la Figure 4.2c-d, une périodicité peut être observée pour le courant et la tension, car la source Fronius régule principalement la tension et ajuste le courant

pour maintenir la stabilité de l'arc, ce qui synchronise la périodicité des deux signaux. Pour ce qui est des données acoustiques, il peut être plus difficile d'observer une périodicité en observant la Figure 4.3.

On constate aux Figures 4.2 et 4.3 que le début et la fin des données sont très différents du reste. Ces données représentent les moments transitoires du démarrage et de l'arrêt de l'arc et du robot de soudage, ce qui explique leur différence avec le reste des signaux. Pour éviter d'augmenter la complexité du problème, les données des soudures ont été conservées entre 0,525 s et 31,9 s, après une synchronisation approximative des données acoustiques sur les données électriques. Le reste des données a été retiré, ce qui devrait normalement aider les modèles à mieux s'ajuster aux données conservées.

La taille de l'ensemble de données utilisé a été un défi important. En effet, le fait de n'avoir que 60 soudures d'aluminium et 76 soudures d'acier n'était pas vraiment suffisant pour entraîner de bons modèles, surtout pour ceux basés sur le DL. Pour remédier à cette situation, chaque soudure a été séparée en plusieurs fractions de soudures pour être en mesure de fournir un plus grand nombre d'instances aux modèles. Les segments de soudures utilisés pouvaient varier entre 0,05 s et 1 s en fonction des exécutions des algorithmes, ce qui signifie qu'une exécution avec des intervalles de 0,25 s avait donc des données de 10 000 points si elles étaient de nature électrique ou 12 000 points si elles étaient acoustiques, ce qui permettait d'avoir environ 41 pulsions (voir la Figure 4.2) dans cet intervalle. En faisant cette séparation des données, on considère l'hypothèse d'auto-similarité qui suppose que chaque fraction contient les mêmes caractéristiques que la soudure complète. Cependant, il est important de reconnaître que des défauts ponctuels peuvent apparaître dans certaines fractions de soudures, ce qui pourrait nuire à la performance des modèles. Le fait de reconnaître cette limitation permet de mieux contextualiser les résultats, notamment lorsque ceux-ci sont moins fiables. Cette

approche devrait donc permettre aux modèles d'obtenir plus de données sans compromettre la qualité de celles-ci, et ainsi les aider à obtenir de meilleures performances.

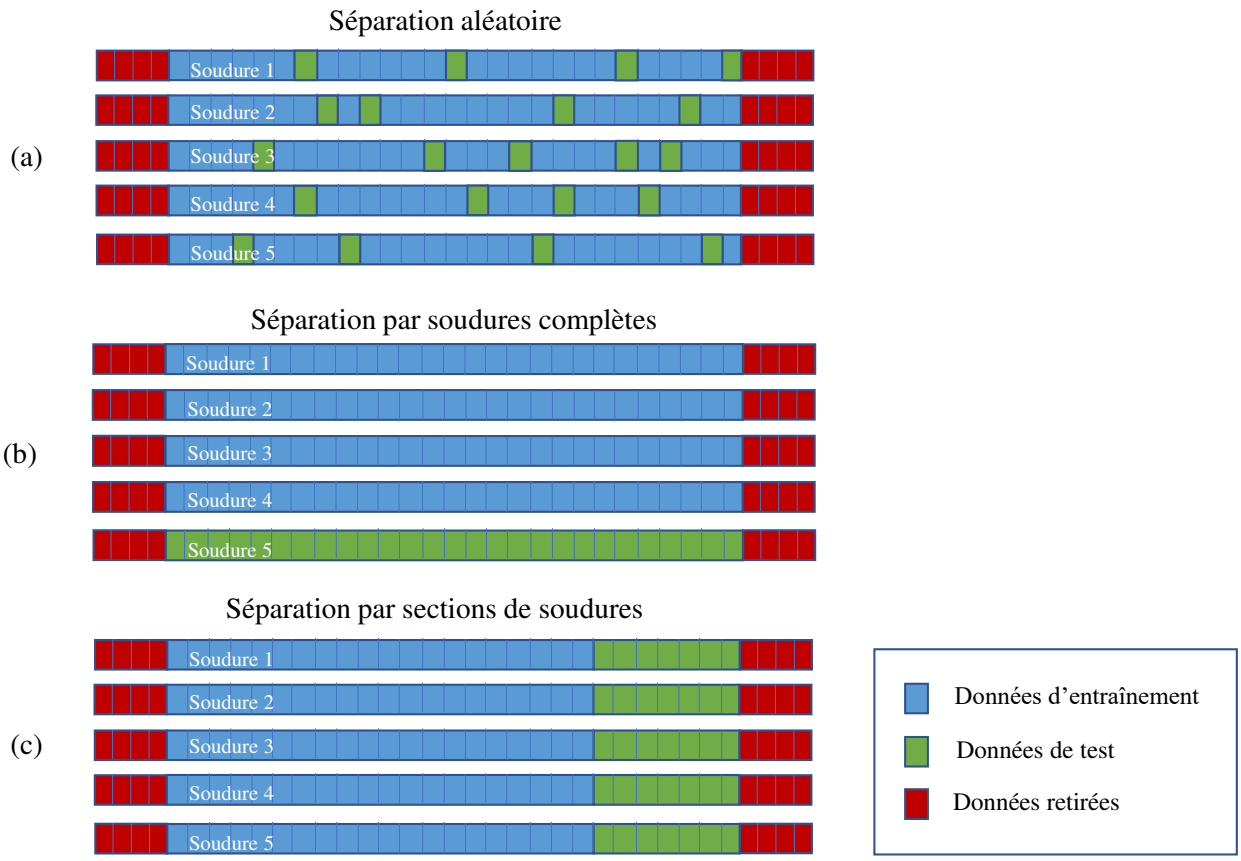


FIGURE 4.4 : Les différentes manières de séparer nos données d'entraînement et de test.

Pour ce qui est de la division entre les données d'entraînement et les données de test, comme on peut le voir dans la Figure 4.4, plusieurs méthodes ont été testés préliminairement avant de choisir celle qui a été considérée comme la plus adaptée pour le contexte actuel. Le but était de trouver une méthode fonctionnelle avec le moins de biais possible pour ainsi avoir des résultats les plus fiables possibles.

Séparation aléatoire : Cette méthode consiste à choisir un pourcentage des données pour l'ensemble de test et de prendre ces données aléatoirement parmi tous les segments de chacune des soudures, sans tenir compte des classes d'anomalies. Bien que cela puisse diversifier les données de test, cela peut faire en sorte que les données soient trop similaires aux données d'entraînement. En effet, les segments adjacents pourraient avoir de plus grandes similarités que s'ils étaient plus espacés dans la soudure.

Séparation par soudures complètes : Cette méthode, qui est théoriquement la plus optimale pour obtenir des résultats plus précis, consiste à prendre des soudures complètes pour l'ensemble de test et ainsi s'assurer que les données d'entraînement et de test proviennent de séquences complètement différentes. Bien que cela permette d'avoir des résultats moins biaisés, cela réduit la diversité des données de test et est difficilement applicable dans le contexte où nous avons seulement 60 soudures. En effet, il y a des situations dans les tests effectués où il n'y a qu'une seule soudure par type et sévérité d'anomalie, ce qui rend la séparation des ensembles impossible.

Séparation par sections de soudures : Cette méthode consiste à isoler un pourcentage adjacent de chaque soudure (début, milieu ou fin) et de l'utiliser comme ensemble de test. Cela peut en effet causer un biais comme la première méthode, mais comme les données d'entraînement et de test ont une moins grande proximité à l'intérieur des soudures, il semble plausible que le biais sera moins grand. Il faut aussi prendre en compte que, tout comme la première méthode, on garde une certaine diversité dans les différents ensembles.

4.2 MODÈLES DE FORÊT ALÉATOIRE

Un RF est une méthode d'apprentissage automatique utilisée pour la classification et la régression. Elle crée plusieurs arbres de décision, qui sont des modèles qui divisent les

données en fonction de caractéristiques spécifiques. Chaque arbre est construit à partir d'un échantillon aléatoire des données, ce qui permet de capturer une variété de caractéristiques et de relations. Les résultats des arbres sont ensuite combinés pour améliorer l'exactitude et éviter le surapprentissage. Cette approche est particulièrement efficace pour traiter des ensembles de données complexes avec de nombreuses variables, comme la classification de soudures, où les relations entre les variables peuvent être subtiles (Breiman, 2001).

La bibliothèque Scikit-Learn (Pedregosa *et al.*, 2011), intégrée au langage de programmation Python, a été choisie pour l'implémentation du modèle en raison de son grand nombre d'algorithmes de classification, ce qui inclut les RF. Une autre librairie importante qui a été utilisée est TSFresh qui a permis d'extraire des caractéristiques à partir des signaux électriques et acoustiques. Cette étape d'extraction peut demander un temps d'exécution non négligeable.

Pour entraîner le modèle de RF, plusieurs hyperparamètres et configurations d'entraînement ont été testés afin de comparer les performances de chaque variante et ainsi obtenir un modèle plus précis. Les hyperparamètres et configurations qui ont été testés sont les suivants :

Type de classe : Cela indique la manière de séparer les classes, tel que vu au début du chapitre actuel. Seulement deux méthodes de séparation de classes ont été utilisées : la classification basée sur la procédure et la classification combinée avec diagnostic. Il y a donc le cas où on classifie les références contre les anomalies et le cas où on a les réussites contre les différentes anomalies. La méthode de classification basée sur les résultats des tests n'a pas été utilisée en raison du fait que le projet recherchait spécifiquement une certaine explicabilité sur les causes des anomalies.

Nombre d'arbres : Cela indique le nombre d'arbres dans la forêt. Un plus grand nombre peut potentiellement augmenter l'exactitude, mais aussi augmenter le temps d'exécution. Les différentes valeurs testées sont 100, 250 et 500 pour essayer de trouver le meilleur compromis. Ces valeurs ont été choisies car 100 était le nombre par défaut et que le

temps d'exécution ne semblait pas trop long, ce qui fait que l'évaluation de plus petites valeurs semblait peu pertinente et que des valeurs plus grandes ont été testées.

Nombre de points par segment : Cela indique le nombre de points par segment de soudures, ce qui représente le temps réel de chaque segment. Il semblerait logique qu'un grand nombre de points par segment donne plus d'information par segment, mais cela résulterait en un moins grand nombre de segments indépendants. Un plus petit nombre de points par segment ferait l'inverse, moins d'information par segment, mais un plus grand nombre de segments au total. Les valeurs 2000, 10 000, 20 000, 1567 et 2753 ont donc été testées pour identifier le meilleur compromis. Les valeurs 1567 et 2753, toutes deux des nombres premiers, ont été choisies afin de limiter l'apparition de motifs répétitifs ou prévisibles dans les segments. Cela réduit la probabilité que la période de segmentation soit un multiple de celle des pulsations, ce qui permet ainsi d'éviter que celles-ci apparaissent toujours aux mêmes positions dans les segments. Cela pourrait ainsi diminuer le risque que les prédictions soient influencées par une synchronisation artificielle des pulsations.

Source de données : Une seule source de données, soit électrique ou acoustique, a été ciblée spécifiquement afin de comparer les performances individuelles de chacune. Cette approche est adoptée, plutôt que d'utiliser les deux sources en même temps, afin de limiter l'investissement en capteurs, tout en permettant de comparer les performances de ces sources de données.

Réduction de dimensionnalité : Cela indique la méthode de réduction de dimensionnalité qui a été utilisé. Les différentes méthodes utilisées sont le Test du khi carré (χ^2), le PCA ainsi qu'une autre méthode où on ajoute une variable aléatoire dans les données pour ensuite exclure les variables que le RF considère comme moins importantes que celle qui est aléatoire.

Le PCA est utilisé pour réduire la dimensionnalité des données en combinant les variables initiales. Il génère un plus petit nombre de nouvelles variables, appelées composantes principales, qui résument au mieux l'information contenue dans les variables d'origine. En théorie, ces composantes principales sont parfaitement orthogonales entre elles, ce qui signifie qu'elles sont statistiquement indépendantes, ce qui permet ainsi de mieux différencier les données et de réduire la redondance. Par contre, il faut mentionner que, comme les caractéristiques originales ne sont pas gardées pour la classification, une bonne partie de l'explicabilité peut être perdue en utilisant cette technique (Abdi & Williams, 2010).

Le test du χ^2 permet de sélectionner les variables les plus importantes en mesurant leur lien statistique avec les différentes classes, et en retirant celles qui ont moins d'influence sur celle-ci. Contrairement à une approche où une variable aléatoire est ajoutée pour établir un seuil de pertinence, le test du χ^2 se base sur l'importance statistique des variables, ce qui permet de créer un classement des variables en fonction de leur contribution à la distinction de la classe. Cette méthode permet de garder les variables les plus pertinentes pour le modèle, tout en offrant une évaluation plus systématique de leur importance (Liu & Setiono, 1995).

Séparation des données : Cela représente les méthodes qui sont illustrées dans la Figure 4.4. La méthode de séparation par soudures complètes a rapidement été abandonnée en raison du nombre de soudure trop bas. Les méthodes de séparation aléatoire et par sections de soudures sont celles qui ont été testées avec 30 % de données pour les tests. De plus, pour augmenter l'échantillonnage statistique, ce qui devrait permettre de réduire les biais, chacune des méthodes a été testée trois fois. Pour la méthode aléatoire, l'état aléatoire (random state) a été utilisé avec trois valeurs différentes, mais constantes, ce qui a permis de tester chaque configuration trois fois pour ainsi obtenir des résultats

moyennés. Pour la méthode par sections de soudure, les données de tests ont été prises à trois endroits différents : au début de la soudure, au milieu de la soudure ou à la fin de la soudure, ce qui a permis d'amener, comme la méthode aléatoire, trois résultats différents par configuration. Les métriques de performance rapportées à la section 5.1 sont les moyennes de ces résultats.

Ensemble de données : Cela indique si l'ensemble de données est celui avec des soudures d'aluminium ou celui avec des soudures d'acier. En effet, après avoir fait les tests sur les soudures d'aluminium, les soudures d'acier ont été testées pour comparer les différents résultats.

4.3 MODÈLES DE CNN

Un CNN est un type de modèle d'apprentissage profond qui est efficace pour extraire automatiquement des motifs locaux dans des données structurées. Ils sont surtout connus pour faire du traitement d'images, mais ils sont aussi très utiles pour analyser des signaux bruts, tels que des signaux électriques ou acoustiques. Pour cela, ils captent des caractéristiques locales grâce à des noyaux appliqués successivement. Ils permettent ainsi d'éviter une phase manuelle de sélection de caractéristiques tout en conservant la structure des données (LeCun & Bengio, 1998). Dans ce projet, les modèles de CNN ont été développés en utilisant les bibliothèques Keras et TensorFlow qui sont disponibles dans le langage Python (Abadi *et al.*, 2016). Cela a permis de faciliter la conception et l'optimisation de ces modèles.

Contrairement au modèle de RF qui utilise des caractéristiques extraites par TSFresh, le CNN traite directement les données brutes. Grâce à sa capacité d'extraction automatique des caractéristiques, le DL ne nécessite pas d'extraction manuelle des caractéristiques, ce qui est particulièrement avantageux pour l'analyse de signaux bruts comme les données électriques et acoustiques de ce projet (LeCun *et al.*, 2015).

Dans certains contextes, des CNN préentraînés peuvent être réentraînés sur de nouvelles données pour potentiellement obtenir de meilleurs résultats et être plus rapides à entraîner qu'un CNN conçu et entraîné à partir de zéro. Toutefois, les modèles préentraînés les plus courants sont conçus pour des images 2D, avec des convolutions typiquement de taille 3x3 ou 5x5, et reposent sur des entrées RGB, ce qui les rend inadaptés à nos données temporelles qui ont un ou deux canaux. Bien qu'il existe aussi des modèles préentraînés spécifiquement conçus pour les séries temporelles, ces derniers sont moins nombreux et moins populaires que ceux en 2D, ce qui limite leur accessibilité. Il a donc été choisi de concevoir une architecture manuellement afin de garder un contrôle total sur la structure du CNN et de l'adapter à notre problème.

Pour la conception de l'architecture de CNN, deux modèles différents ont été conçus pour les entraînements avec les données acoustiques et électriques. Le modèle pour les données acoustiques qui ont un seul canal a été conçu avec des convolutions 1D, tandis que celui pour les données électriques a été conçu avec des convolutions 2D pour pouvoir interagir avec les deux canaux en même temps. L'architecture des modèles a été déterminée à partir de principes connus de conception de CNN, puis ajustée à la suite de plusieurs tests expérimentaux réalisés lors de la phase de conception initiale.

4.3.1 CNN POUR LES DONNÉES ACOUSTIQUES

```
def get_model_1d(num_classes, default_strides=True, regularizer_value=0.01,
                 nb_channels=1, input_length=2000, activation='relu'):
    model = Sequential()

    strides = 1

    if input_length <= 2400 or default_strides:
        strides = 1
    elif input_length <= 12000:
```

```
        strides = 4
    elif input_length <= 24000:
        strides = 7

    # Première couche de convolution
    model.add(Conv1D(32, 50, strides=strides, activation=activation,
                     input_shape=(input_length, nb_channels),
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(AveragePooling1D(5))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Deuxième couche de convolution
    model.add(Conv1D(64, 20, activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(AveragePooling1D(5))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Troisième couche de convolution
    model.add(Conv1D(128, 10, activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(MaxPooling1D(5))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Quatrième couche de convolution
    model.add(Conv1D(256, 5, activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(MaxPooling1D(5))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Couche de flatten et couches denses
    model.add(Flatten())
    model.add(Dense(512, activation=activation,
                    kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(Dropout(0.5))
    model.add(Dense(num_classes, activation='softmax',
                    kernel_regularizer=regularizers.l2(regularizer_value)))

    model.compile(optimizer=Adam(learning_rate=0.0001),
                  loss='sparse_categorical_crossentropy', metrics=['accuracy'])

    return model
```

Le code précédent démontre l'architecture qui a été conçue pour traiter les données acoustiques. Le modèle est principalement constitué de couches de convolution qui sont suivies par des couches de sous-échantillonnage (pooling), de normalisation par lots et de masquage aléatoire de neurones (dropout).

Comme il peut être constaté, le nombre de pas (stride) peut être plus élevé lorsque les données testées sont plus longues. Le nombre de pas est appliqué à la première convolution, ce qui permet une plus grand réduction des données, et donc, du nombre de paramètres dans le modèle.

Pour ce qui est de la largeur des filtres, la première couche de convolution utilise un filtre de 50 points, ce qui équivaut à 1,04 ms de données acoustiques et à 1,25 ms de données électriques. Ce choix de longueur de filtre permet de capturer des motifs de taille intermédiaire dans les données. Bien que la taille des filtres diminue dans les couches suivantes (20, 10 et 5 points), ces couches exploitent les caractéristiques fines extraites précédemment afin de capturer des motifs plus larges et plus complexes dans le signal. Les données sont ensuite principalement réduites grâce aux couches de sous-échantillonnage qui sont utilisées après chaque couche de convolution. Ces couches permettent de réduire la taille des données en sélectionnant les valeurs maximales ou moyennes dans des régions locales. Cela permet de conserver les caractéristiques essentielles en filtrant les détails moins importants, ce qui contribue à une meilleure généralisation du modèle.

Plusieurs techniques ont été ajoutées au modèle pour aider à éviter le surapprentissage observé lors des entraînements préliminaires. Le masquage aléatoire de neurones est utilisé pour prévenir le surapprentissage en désactivant de manière aléatoire certains neurones pendant l'entraînement, ce qui permet au modèle d'éviter d'être trop dépendant à certains neurones. La normalisation par lots contribue à stabiliser et accélérer l'entraînement en normalisant les

valeurs des sorties des fonctions d'activation. La régularisation L2 est appliquée dans chaque couche de convolution pour réduire la valeur des poids qui sont trop élevés, ce qui réduit le risque de surapprentissage sur les données d'entraînement.

Après les couches de convolution, un aplatissement des données (couche de flatten) est effectué pour transformer les données en un vecteur à une dimension. Cela permet donc à une couche dense d'effectuer la classification grâce à une fonction softmax qui produit des probabilités pour chaque classe, permettant ainsi au modèle de faire des prédictions de classification.

4.3.2 CNN POUR LES DONNÉES ÉLECTRIQUES

```
def get_model_2d(num_classes, default_strides=True, regularizer_value=0.01,
                 nb_channels=1, input_shape=(2, 2000), activation='relu'):
    model = Sequential()

    strides = (1, 1)

    if input_shape[1] <= 2400 or default_strides:
        strides = (1, 1)
    elif input_shape[1] <= 12000:
        strides = (1, 4)
    elif input_shape[1] <= 24000:
        strides = (1, 7)

    # Première couche de convolution
    model.add(Conv2D(32, (2, 50), strides=strides, activation=activation,
                     input_shape=input_shape + (nb_channels,),
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(AveragePooling2D((1, 5)))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Deuxième couche de convolution
    model.add(Conv2D(64, (1, 20), activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(AveragePooling2D((1, 5)))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())
```

```
    # Troisième couche de convolution
    model.add(Conv2D(128, (1, 10), activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(MaxPooling2D((1, 5)))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Quatrième couche de convolution
    model.add(Conv2D(256, (1, 5), activation=activation,
                     kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(MaxPooling2D((1, 5)))
    model.add(Dropout(0.25))
    model.add(BatchNormalization())

    # Couche de flatten et couches denses
    model.add(Flatten())
    model.add(Dense(512, activation=activation,
                    kernel_regularizer=regularizers.l2(regularizer_value)))
    model.add(Dropout(0.5))
    model.add(Dense(num_classes, activation='softmax',
              kernel_regularizer=regularizers.l2(regularizer_value)))

    model.compile(optimizer=Adam(learning_rate=0.0001),
                  loss='sparse_categorical_crossentropy', metrics=['accuracy'])

    return model
```

Le CNN conçu pour les données électriques est presque identique à celui qui sert à classifier les données acoustiques. Bien que toutes les opérations de convolutions et d'échantillonnage soient en deux dimensions, la plupart d'entre elles sont techniquement en une dimension car elles ont une hauteur de 1. La seule opération qui a une hauteur de 2 est la première convolution, ce qui permet de capturer des caractéristiques pour la tension et le courant dans la même opération, et donc, de combiner leurs caractéristiques. Des tests préliminaires ont montré que cette approche semblait aider à obtenir une meilleure convergence du modèle. Elle s'est avérée beaucoup plus efficace qu'une combinaison à la fin du réseau, probablement grâce à la capacité à extraire et intégrer les caractéristiques communes dès le début du processus, ce qui faciliterait ainsi l'apprentissage des relations entre les deux séries.

Tout comme le RF, le CNN a été testé avec plusieurs variations d'hyperparamètres et autres configurations d'entraînement dans le but d'optimiser la classification. Les hyperparamètres et configurations d'entraînement sont les suivants :

Type de classe : Les trois méthodes de séparation des classes ont été testées dans le CNN, même si la méthode basée uniquement sur les résultats des tests est moins utile pour les objectifs du projet.

Nombre de points par segment : Cela indique le nombre de points par segment de soudures. Les valeurs 2000, 10 000, 20 000, 1567 et 2753 ont été testées pour essayer de trouver le meilleur compromis, tel qu'expliqué dans la section 4.2.

Source de données : Comme pour le RF, pour limiter l'investissement de capteurs, les sources électriques et acoustiques ont été testées séparément dans le but de comparer leurs performances individuelles.

Séparation des données : Cela représente les méthodes qui sont illustrées dans la Figure 4.4. Tout comme lors du RF, la méthode de séparation par soudures complètes a été presque pas du tout utilisée en raison du nombre de soudure trop bas. Les méthodes de séparation aléatoire et par sections de soudures sont encore celles qui ont été testées, mais maintenant avec 20 % de données pour les tests et 20 % pour la validation. Dans le contexte des algorithmes de DL, l'ensemble de validation est utilisé après chaque passage complet sur les données d'entraînement pour ajuster les hyperparamètres et évaluer le modèle en cours d'entraînement, afin de prévenir le surapprentissage. Les données de test, quant à elles, sont utilisées seulement à la fin de l'entraînement pour obtenir une évaluation finale du modèle, comme c'était le cas avec les algorithmes de ML classique. Encore une fois, chacune des méthodes ont été testées trois fois pour tenter de réduire les biais. Pour la méthode aléatoire, l'état aléatoire (random state) a encore été utilisé avec trois valeurs différentes et constantes. Pour la méthode par

sections de soudure, les données ont été séparées par quintiles de soudures pour avoir des sections de 20 % chacune. Il y a donc eu trois manières de prendre les données des quintiles pour les utiliser dans les ensembles de données. Les trois méthodes de séparation sont les suivantes : utiliser le quintile 1 pour le test, le quintile 3 pour la validation, et le reste pour l'entraînement ; utiliser le quintile 2 pour le test, le quintile 4 pour la validation, et le reste pour l'entraînement ; ou utiliser le quintile 3 pour le test, le quintile 5 pour la validation, et le reste pour l'entraînement. Cela permet donc d'avoir trois résultats différents comme la méthode aléatoire. De plus, les quintiles de validation et de test sont toujours espacés par un quintile d'entraînement. Cela permet d'éviter d'avoir un intervalle de 40 % sans données d'entraînement, ce qui devrait permettre de garder de la diversité dans les données d'entraînement. De plus, il est important de mentionner que les ensembles de données d'entraînement et de test sont différents entre le RF et le CNN, même pour la méthode aléatoire, ce qui fait que la comparaison entre les deux types de modèles ne sera pas directe. Bien que les ensembles d'entraînement soient plus petits dans les algorithmes de DL, l'utilisation de l'ensemble de validation permet d'ajuster les hyperparamètres pour améliorer les performances.

Méthode d'échantillonnage de l'ensemble de données : Cela représente les méthodes d'échantillonnage qui ont été utilisés, soit le suréchantillonnage et le sous-échantillonnage. Le suréchantillonnage consiste à augmenter artificiellement le nombre d'exemples dans les classes minoritaires, par duplication, où les données existantes sont simplement copiées pour augmenter leur nombre, ou par génération synthétique, où de nouvelles données sont créées en combinant ou en modifiant les caractéristiques des données existantes, pour ainsi compenser leur nombre inférieur aux autres classes. Le sous-échantillonnage consiste à réduire le nombre d'exemples dans les classes majoritaires afin d'équilibrer la distribution des classes dans l'ensemble d'apprentissage. Il est important de mentionner que les données de test ne sont jamais modifiées par

ces techniques, afin d'obtenir une évaluation non biaisée des performances du modèle. Le modèle a donc pu être entraîné avec ces méthodes d'échantillonnage ou avec son ensemble de données non altéré pour pouvoir amener des comparaisons.

Fonction d'activation : Différentes fonctions d'activation ont été testées afin de comparer l'efficacité de chacune d'entre elles. Les fonctions sont les suivantes : ReLU, ELU, linéaire, sigmoïde, Tanh et Swish (Dubey *et al.*, 2022).

Normalisation des données : La normalisation des données a été testée pour évaluer son impact sur la performance du modèle. Il a donc été testé avec et sans normalisation, afin de comparer les résultats obtenus. La normalisation vise à standardiser l'échelle des caractéristiques, ce qui peut potentiellement améliorer l'efficacité du modèle.

Séparation des classes positives et négatives : Tel que vu précédemment dans la Figure 3.1, certaines classes (hors-position, longueur d'arc et débit de gaz de protection) peuvent être divisées en sous-classes positives et négatives. Cela offre donc la possibilité de regrouper les classes positives et négatives sous une même classe ou de les classer séparément. Il y a deux hypothèse derrière cela : la première est que les classes positives seraient assez différentes des classes négatives, ce qui permettrait une meilleure classification lorsqu'elles sont divisées ; la seconde est que les classes négatives et positives seraient plutôt similaires, ce qui les rendraient plus faciles à classifier de manière groupée.

Ensemble de données : Tout comme le RF, le CNN a été testé sur l'ensemble de données de soudures d'acier après les tests sur l'aluminium.

Apprentissage par transfert : L'apprentissage par transfert a été testé pour tenter d'améliorer la classification des modèles. Bien qu'il a été mentionné précédemment que des modèles préentraînés n'ont pas été utilisés, il a été possible d'effectuer l'apprentissage par transfert à partir des modèles entraînés lors du projet. En effet, les modèles

entraînés sur l'aluminium ont pu être réentraînés sur l'acier et inversement. De plus, dans un CNN, il faut geler un certain nombre de couches au début du réseau pour ne garder que quelques couches entraînables à la fin pour être en mesure de faire de l'apprentissage par transfert. Il y a donc plusieurs configurations de nombre de couches entraînables qui ont été testées afin de comparer leur efficacité : 4 couches, 8 couches, 12 couches et 16 couches. Cet apprentissage par transfert pourrait être bénéfique dans notre contexte de données limitées, car il pourrait permettre d'exploiter les connaissances acquises sur nos deux ensembles de données qui sont similaires.

Nombre de pas (strides) : Tel que vu précédemment dans le code, le nombre de pas peut être augmenté lorsque les données sont plus longues. Cela a pour but de réduire le coût computationnel, et donc, d'accélérer l'entraînement du modèle. Les valeurs possibles sont 1, 4 et 7 en fonction de la longueur des segments de soudure.

4.3.3 MODÈLE ADAPTÉ À DES SCALOGRAMMES

Un autre modèle de CNN a aussi été testé pour vérifier son efficacité de classification. Ce modèle utilise des scalogrammes, qui sont des représentations visuelles des signaux dans le domaine temps-fréquence. L'utilisation de scalogrammes permet de capturer des caractéristiques dynamiques et locales des signaux, ce qui pourrait aider à l'identification des motifs pertinents pour la classification.

Comme ces scalogrammes donnent des images qui ont une hauteur de 264 et une largeur de 2000, le CNN adapté à ces données est pratiquement identique à celui pour les données électriques brutes. La seule différence notable est que la première convolution et la deuxième ont des hauteurs respectives de 5 et 2, ce qui devrait aider à extraire des caractéristiques dans des zones qui ont une certaine hauteur.

Étant donnée la lourdeur que cela implique en terme de puissance de calcul, très peu de configurations différentes ont été testées. Les scalogrammes ont été testés en nuances de gris ou en couleur pour voir si cela impactait la classification. De la manière dont ils ont été générés, un seul signal pouvait être transformé en scalogramme. La tension et le courant ont donc été testés séparément, en plus de l'amplitude acoustique ainsi que la combinaison des signaux électriques pour obtenir une résistance effective instantanée. Le modèle a aussi été testé avec les ensembles de données d'aluminium et d'acier pour vérifier la transférabilité. De plus, des états aléatoires (random state) ont été utilisés pour tenter d'avoir des résultats moins biaisés. Il faut aussi mentionner que pour des raisons de temps et de complexité de code, seule la méthode de séparation de données aléatoire a été utilisée.

CHAPITRE V
RÉSULTATS ET ANALYSE

Le chapitre suivant présente une analyse approfondie des résultats obtenus lors des tests des modèles de classification. L'objectif est de comparer les performances des modèles de RF et de CNN, en évaluant leur capacité à classifier les soudures selon les différents paramètres testés. Les implications de ces résultats pour les applications industrielles sont également discutées. De plus, une analyse détaillée des résultats obtenus suite aux changements dans la configuration de soudage sera présentée, afin d'évaluer l'impact de ces modifications sur les performances de classification.

Les différents modèles ont été entraînés en itération de plusieurs tests avec plusieurs variations d'hyperparamètres et de configurations. Les résultats présentés représenteront des moyennes d'exactitude pour chaque variante de configuration, mais toutes les autres variantes seront incluses dans ces moyennes. En raison de cela, les exactitudes peuvent être plus basses que l'optimal.

5.1 RÉSULTATS DE LA FORÊT ALÉATOIRE

Plusieurs itérations ont été faites pour obtenir les différents résultats du RF. Chacune d'entre elle a servi à tester différentes combinaisons d'hyper-paramètres et de configurations, sans avoir toutes les combinaisons possibles en même temps.

La première itération de tests a été faite avec l'ensemble de données d'aluminium. Seule les méthodes de séparation aléatoires et par soudures complètes ont été testées parce que la méthode par section de soudures n'avait pas été implémentée à ce moment dans le projet. De plus, seule la méthode de réduction de dimensionnalité avec la variable aléatoire a été utilisée

puisque les autres n'étaient pas encore implémentées. Il faut aussi prendre en compte que les classes positives et négatives ont été regroupées sous la même classe lors de ces tests. 576 tests individuels ont été faits et les résultats pour chaque variante de configuration sont les suivants :

TABLEAU 5.1 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification basée sur la procédure	0,739
Classification combinée avec diagnostic	0,733

TABLEAU 5.2 : Exactitude selon le nombre d'arbres

Nombre d'arbres	Exactitude
100 arbres	0,733
250 arbres	0,735
500 arbres	0,738

TABLEAU 5.3 : Exactitude selon le nombre de points par segment et la source de données

Nombre de points	Exactitude (électrique)	Exactitude (acoustique)
1567 points	0,750	0,562
2000 points	0,758	0,570
2753 points	0,769	0,579
10 000 points	0,811	0,624
20 000 points	0,882	0,822
40 000 points	0,876	0,825

TABLEAU 5.4 : Exactitude selon la source de données

Source de données	Exactitude
Données électriques	0,808
Données acoustiques	0,663

Le Tableau 5.1 montre une classification légèrement meilleure lorsque les classes sont basées sur la procédure. Le Tableau 5.2 suggère que le nombre d'arbres n'aurait pas d'impact considérable sur la classification lorsque fixé à des valeurs supérieures à la valeur par défaut de 100. Le Tableau 5.3 semble montrer que les nombres premiers n'ont pas vraiment d'impact sur

TABLEAU 5.5 : Impact de la réduction de dimensionnalité sur l'exactitude

Réduction de dimensionnalité	Exactitude
Avant suppression	0,731
Après suppression	0,741

TABLEAU 5.6 : Exactitude selon la méthode de séparation des données

Séparation des données	Exactitude
Aléatoire	0,776
Soudures différentes	0,612

la classification par rapport aux segments de 2000 points qui sont les plus ressemblants en terme de longueur. Aussi, de plus grands segments, de l'ordre d'une seconde ou moins, semblent donner de meilleurs résultats de classification, et ce, autant pour les données électriques que les données acoustiques. Le Tableau 5.4 montre que les données électriques ont obtenu une meilleure classification que les données acoustiques. Selon le Tableau 5.5, la réduction de dimensionnalité avec une variable aléatoire permettrait d'améliorer légèrement la classification. Le Tableau 5.6 montre une certaine différence entre les méthodes de séparation de données, ce qui indique que, tel que supposé précédemment, il est possible que les données séparées aléatoirement au travers des soudures amènent un certain biais dans les modèles, ce qui amène à des prédictions trop optimistes.

La deuxième itération de tests a aussi été faite seulement avec l'ensemble de données d'aluminium. Les tests de réduction de dimensionnalité et de classes positives et négatives ont été faits de la même manière que précédemment. Par contre, les méthodes de séparation des ensembles de données d'entraînement et de test ont tous été testées. De plus, le nombre d'arbres a été de 100 pour tous les tests étant donné qu'il ne semblait pas faire de différence considérable. Les nombres de segments ont été limités aux nombres de 2000, 10 000 et 20 000, pour pouvoir faire des comparaisons avec le CNN qui ne pouvait pas être entraîné sur des

données plus longues en raison de limites de mémoire vidéo. Au total, 120 tests ont été faits et les résultats sont les suivants :

TABLEAU 5.7 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,717
Classification basée sur la procédure	0,692

TABLEAU 5.8 : Exactitude selon le nombre de points par segment et la source de données

Nombre de points	Exactitude (électrique)	Exactitude (acoustique)
2000 points	0,754	0,560
10 000 points	0,818	0,611
20 000 points	0,840	0,643

TABLEAU 5.9 : Exactitude selon la source de données

Source de données	Exactitude
Données électriques	0,804
Données acoustiques	0,605

TABLEAU 5.10 : Impact de la réduction de dimensionnalité sur l'exactitude

Réduction de dimensionnalité	Exactitude
Avant suppression	0,702
Après suppression	0,707

TABLEAU 5.11 : Exactitude selon la méthode de séparation des données

Séparation des données	Exactitude
Aléatoire	0,743
Section de soudure	0,717
Soudures différentes	0,577

Le Tableau 5.7 montre que la classification combinée avec diagnostic a une exactitude légèrement supérieure à celle basée sur la procédure, ce qui contredit les tests précédents. Le

Tableau 5.8 révèle que les segments de 20 000 points offrent la meilleure exactitude, tandis que les segments plus courts de 10 000 et 2000 points, montrent des exactitudes moindres, et ce, pour les deux sources de données. De plus, il est important de remarquer que l'exactitude a considérablement baissé par rapport à la première itération pour les données de 20 000 points, et ce, en particulier pour les données acoustiques [1]. Le Tableau 5.9 indique que les données électriques sont encore une fois meilleures que les données acoustiques en termes d'exactitude dans la classification. Selon le Tableau 5.10, la réduction de dimensionnalité avec une variable aléatoire, améliore encore très légèrement l'exactitude. Enfin, le Tableau 5.11 montre que la méthode de séparation aléatoire des données offre encore une exactitude moyenne supérieure, tandis que la méthode par section de soudure semble être un bon compromis entre les autres méthodes de séparation.

La troisième itération a été plus courte que les autres avec seulement 24 tests au total pour l'ensemble de données d'aluminium. Le but était d'effectuer peu de tests dans de bonnes conditions pour les comparaisons. Pour cela, seule la méthode de séparation par section de soudure a été utilisée en raison du compromis mentionné précédemment. Le nombre d'arbres est toujours à 100 et la longueur des segments a été limité à seulement 10 000 pour pouvoir faire des comparaisons avec le CNN, les explications seront détaillées dans la section 5.2. De plus, les classes positives et négatives ont été séparées pour tester la classification de manière plus détaillée, ce qui permet de faire une classification qui serait plus représentative de ce qui pourrait se faire en industrie.

Le Tableau 5.12 montre que la classification combinée avec diagnostic atteint une exactitude supérieure par rapport à celle basée sur la procédure. Le Tableau 5.13 indique que les données électriques ont obtenu une meilleure exactitude que les données acoustiques, ce

1. La cause exacte de cette baisse d'exactitude n'a pas été identifiée, mais il est possible qu'une modification non documentée du code ou des données ait corrigé un biais qui aurait initialement donné de meilleurs résultats.

TABLEAU 5.12 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,722
Classification basée sur la procédure	0,659

TABLEAU 5.13 : Exactitude selon la source de données

Source de données	Exactitude
Données électriques	0,793
Données acoustiques	0,588

TABLEAU 5.14 : Impact de la réduction de dimensionnalité sur la exactitude

Réduction de dimensionnalité	Exactitude
Avant suppression	0,689
Après suppression	0,692

TABLEAU 5.15 : Exactitude selon le type de classe et source de données

Source de données	Type de classe	Exactitude
Données électriques	Classification combinée avec diagnostic	0,794
	Classification basée sur la procédure	0,792
Données acoustiques	Classification combinée avec diagnostic	0,650
	Classification basée sur la procédure	0,526

qui est cohérent avec les itérations précédentes. Le Tableau 5.14 montre, encore une fois, que la réduction de dimensionnalité avec la variable aléatoire améliore de manière négligeable l'exactitude. Enfin, le Tableau 5.15 montre une comparaison des sources de données combinées avec le type de classe. Les données électriques obtiennent une meilleure exactitude, que ce soit avec une ou l'autre des méthodes de classification, par rapport aux données acoustiques. De plus, pour les deux sources de données, la classification combinée avec diagnostic a obtenu des résultats supérieurs à celle basée sur la procédure. Cette comparaison est utile pour mieux comprendre les résultats en fonction des différents objectifs du projet.

Cette troisième itération a été exécutée une deuxième fois, de manière complètement identique, avec l'ensemble de données de soudures d'acier. Voici les résultats des 24 tests :

TABLEAU 5.16 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,733
Classification basée sur la procédure	0,594

TABLEAU 5.17 : Exactitude selon la source de données

Source de données	Exactitude
Données électriques	0,737
Données acoustiques	0,590

TABLEAU 5.18 : Impact de la réduction de dimensionnalité sur l'exactitude

Réduction de dimensionnalité	Exactitude
Avant suppression	0,662
Après suppression	0,665

TABLEAU 5.19 : Exactitude selon le type de classe et source de données

Source de données	Type de classe	Exactitude
Données électriques	Classification combinée avec diagnostic	0,772
	Classification basée sur la procédure	0,701
Données acoustiques	Classification combinée avec diagnostic	0,694
	Classification basée sur la procédure	0,486

Le Tableau 5.16 montre que la classification combinée avec diagnostic obtient une exactitude nettement supérieure à celle basée sur la procédure. Le Tableau 5.17 indique que les données électriques ont obtenu une meilleure exactitude par rapport aux données acoustiques. Le Tableau 5.18 démontre que la réduction de dimensionnalité avec une variable aléatoire permet d'améliorer très légèrement l'exactitude. Enfin, le Tableau 5.19 montre une comparaison des sources de données combinées avec le type de classe. Les données électriques

ont obtenu une meilleure exactitude, avec les deux types de classes, par rapport aux données acoustiques. Encore une fois, cette comparaison a pour but d'aider à mieux comprendre les résultats en fonction des différents objectifs du projet. Dans tous les cas, les résultats ont été très similaires pour l'ensemble de données sur l'acier par rapport à celui de l'aluminium.

Une dernière itération a été faite dans le but de comparer seulement les différentes méthodes de réduction de dimensionnalité, puisqu'elles n'avaient pas été testées précédemment. Donc, les méthodes de PCA et de χ^2 ont pu être comparé avec la méthode utilisant une variable aléatoire. La méthode de séparation des données utilisée a été celle des portions de soudures, la longueur des segments a été de 10 000 (12 000 pour les données acoustiques) points et le nombre d'arbres est resté constant à 100. Les données électriques et acoustiques ont été testées ainsi que les différentes classes basées sur la procédure et combinées avec diagnostic. De plus, les tests de χ^2 et PCA ont été faits avec plusieurs variantes. En effet, ils ont été chacun testés de trois manières différentes pour diviser le nombre de caractéristiques par 2, 4 et 8. Au total, 192 tests ont été faits sur les deux ensembles de données.

TABLEAU 5.20 : Comparaison des performances de classification pour les données électriques.

Méthode de réduction de dimensionnalité	Exactitude
Sans réduction	0,764
Variable aléatoire	0,765
χ^2 (caractéristiques / 2)	0,763
χ^2 (caractéristiques / 4)	0,708
χ^2 (caractéristiques / 8)	0,658
PCA (caractéristiques / 2)	0,665
PCA (caractéristiques / 4)	0,653
PCA (caractéristiques / 8)	0,646

Les résultats démontrent que la méthode avec la variable aléatoire permet de légèrement améliorer l'exactitude pour les données acoustiques, tandis que l'amélioration est très négligeable pour les données électriques. Pour le χ^2, il y a eu une amélioration négligeable

TABLEAU 5.21 : Comparaison des performances de classification pour les données acoustiques.

Méthode de réduction de dimensionnalité	Exactitude
Sans réduction	0,587
Variable aléatoire	0,595
χ^2 (caractéristiques / 2)	0,590
χ^2 (caractéristiques / 4)	0,541
χ^2 (caractéristiques / 8)	0,485
PCA (caractéristiques / 2)	0,545
PCA (caractéristiques / 4)	0,523
PCA (caractéristiques / 8)	0,490

pour les données acoustiques lorsque le nombre de caractéristique a été divisé par 2. Dans les autres tests avec le χ^2, l'exactitude devient de moins en moins bonne lorsque le nombre de caractéristique est divisé par 4 et 8. Le PCA a lui aussi obtenu des exactitudes de moins en moins bonnes lorsque les caractéristiques ont été divisés par 2, 4 et 8. En prenant en compte ces résultats, on peut supposer qu'un plus grand nombre de caractéristique aide généralement à obtenir une meilleure classification dans le cas de ce problème complexe car la méthode aléatoire n'a retiré qu'environ 42 % des caractéristiques.

5.2 RÉSULTATS DU CNN

Tout comme le RF, le CNN a été testé avec plusieurs itérations contenant chacune un certain nombre de tests. Ces différents tests permettent de comparer les différents hyperparamètres et configurations.

La première itération a permis de tester un bon nombre de variations de configurations dans le modèle de CNN. Toutes les méthodes d'échantillonnage et fonctions d'activation mentionnées précédemment ont été testées. Les deux sources de données, la présence ou l'absence de normalisation, la séparation ou non des classes positives et négatives, ainsi que les

trois types de classe ont été testés. Pour la séparation des données, seule la méthode aléatoire a été testée étant donné qu'il ne s'agit que d'une première itération. Pour le nombre de points par segment, seul les valeurs de 2000 et 2753 ont été testées pour rendre l'exécution du code plus légère vu le grand nombre de configurations qui sont déjà testés. Seul l'ensemble de données de l'aluminium a été testé dans cette itération, ce qui fait que l'apprentissage par transfert n'a pas été testé. De plus, le nombre de pas pour la première convolution est toujours resté à 1. Au total, 864 tests différents ont été faits.

TABLEAU 5.22 : Exactitude selon le type d'échantillonnage

Type d'échantillonnage	Exactitude
Aucun	0,647
Suréchantillonnage	0,630
Sous-échantillonnage	0,609

TABLEAU 5.23 : Exactitude selon la fonction d'activation

Activation	Exactitude
elu	0,687
linéaire	0,623
relu	0,748
sigmoïde	0,357
swish	0,696
tanh	0,659

TABLEAU 5.24 : Exactitude selon le nombre de points par segment

Nombre de points par segment	Exactitude
2000 points	0,617
2753 points	0,639

Le Tableau 5.22 montre que les données inchangées ont obtenu la meilleure exactitude, ce qui est meilleur que le suréchantillonnage et le sous-échantillonnage. Le Tableau 5.23 indique que la fonction d'activation "relu" offre la meilleure exactitude parmi les différentes fonctions testées, tandis que la fonction sigmoïde a obtenu l'exactitude la plus basse. Selon le

TABLEAU 5.25 : Exactitude selon la source de données

Type de données	Exactitude
Données électriques	0,657
Données acoustiques	0,600

TABLEAU 5.26 : Impact de la normalisation sur l'exactitude

Utilisation ou non de normalisation	Exactitude
Sans normalisation	0,665
Avec normalisation	0,592

TABLEAU 5.27 : Exactitude selon la séparation des classes positives et négatives ou non

Séparation ou non des classes positives et négatives	Exactitude
Classes positives et négatives séparées	0,611
Classes positives et négatives regroupées	0,646

TABLEAU 5.28 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,609
Classification basée sur les résultats des tests	0,693
Classification basée sur la procédure	0,584

Tableau 5.24, les segments de 2753 points ont une exactitude légèrement supérieure à ceux de 2000 points, mais le but était surtout de tester si le nombre premier pouvait obtenir une moins bonne classification, ce qui indiquerait un possible biais. Le Tableau 5.25 montre que les données électriques ont obtenu une meilleure exactitude que les données acoustiques. Le Tableau 5.26 montre que la normalisation réduit l'exactitude par rapport aux données inchangées. Le Tableau 5.27 indique que regrouper les classes positives et négatives améliore l'exactitude comparativement à leur séparation. Le Tableau 5.28 montre que la classification basée sur les résultats des tests (classification binaire) atteint la meilleure exactitude, surpassant la classification combinée avec diagnostic et celle basée sur la procédure.

Pour la deuxième itération, les méthodes de suréchantillonnage et de sous-échantillonnage n'ont pas été utilisées car les résultats précédents n'étaient pas meilleurs que les données non modifiées. Seule la fonction d'activation ReLU a été utilisée en raison de ses résultats supérieurs aux autres fonctions d'activation. Pour les données électriques, les segments de 2000, 10 000 et 20 000 points ont été testées. Pour les données acoustiques, des segments de longueur 2400, 12 000 et 24 000 ont été testés pour conserver les mêmes durées physiques. La normalisation n'a pas été utilisée puisqu'elle n'améliorait pas l'exactitude. Les classes ont seulement été utilisées dans leur version avec la séparation des classes positives et négatives, étant donné que, même si la classification était moins bonne, cela permet d'augmenter la spécificité de la classification. Pour ce qui est des types de classes, seules les méthodes de classification basée sur la procédure et de classification combinées avec diagnostic ont été utilisées pour mieux répondre aux besoins du projet où l'objectif était de classifier les anomalies selon leur type. Cette itération a seulement été faite avec l'ensemble de données de l'aluminium. Pour cette itération, dans la première convolution du réseau, différents nombres de pas ont été testés : 1, 4 et 7 dépendamment de la longueur des segments de données. Au total, 80 tests différents ont été faits.

TABLEAU 5.29 : Exactitude selon la méthode de séparation des données

Séparation des données	Exactitude
Aléatoire	0,635
Quintiles de soudures	0,567

TABLEAU 5.30 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,712
Classification basée sur la procédure	0,524

Le Tableau 5.29 montre que la séparation aléatoire des données produit une exactitude supérieure comparée aux quintiles de soudures, ce qui est logique selon le fait que les données

TABLEAU 5.31 : Exactitude selon le nombre de pas dans la première convolution et le nombre de points par segment pour les données électriques

	Exactitude selon le nombre de pas		
Nombre de points	**1 pas**	**4 pas**	**7 pas**
2000 points	0,592	-	-
10 000 points	0,462	0,606	-
20 000 points	0,355	-	0,479

TABLEAU 5.32 : Exactitude selon le nombre de pas dans la première convolution et le nombre de points par segment pour les données acoustiques

	Exactitude selon le nombre de pas		
Nombre de points	**1 pas**	**4 pas**	**7 pas**
2400 points	0,757	-	-
12 000 points	0,849	0,792	-
24 000 points	0,627	-	0,660

TABLEAU 5.33 : Exactitude selon la source de données

Source de données	**Exactitude**
Données électriques	0,499
Données acoustiques	0,737

seraient plus biaisées avec cette méthode. Le Tableau 5.30 indique que la classification combinée avec diagnostic a obtenu une exactitude supérieure par rapport à celle basée sur la procédure. Le Tableau 5.31 démontre qu'augmenter le nombre de pas dans la première convolution pour les segments de 10 000 et 20 000 points permet d'augmenter l'exactitude. Les meilleurs modèles électriques ont été ceux avec des segments de 10 000 points et des convolutions initiales de 4 pas. Le Tableau 5.32 montre que pour les données acoustiques, l'augmentation du nombre de pas réduit l'exactitude dans les segments de 12 000 points et augmente légèrement l'exactitude pour les segments de 24 000 points. Les meilleurs modèles acoustiques, qui ont aussi été les meilleurs modèles en général, ont été ceux avec des segments de 12 000 (0,25 s) et des convolutions de 1 pas. Le Tableau 5.33 montre que les

données acoustiques ont obtenu une meilleure exactitude que les données électriques, ce qui est contradictoire avec les tests de la première itération. Il est fort probable que cette différence est dû aux nombreux changements de configuration entre les deux itérations, notamment en raison des segments qui sont désormais plus longs.

Une troisième itération a été faite avec seulement 12 tests avec des conditions qui ont été considérées comme optimales. Aucune méthode d'échantillonnage ou de normalisation n'a été utilisée. Les classes négatives et positives sont restées distinctes. La fonction d'activation ReLU a été la seule utilisée. Les types de classes basées sur la procédure et combinées avec diagnostic sont les deux seuls qui ont été utilisés. Pour la séparation des données, la méthode par sections de soudure avec les quintiles est celle qui a été utilisée. Le nombre de points par segment a été de 10 000 pour les données électriques et 12 000 pour les données acoustiques, étant donné que cela avait donné les meilleurs résultats pour les données acoustiques précédemment. Bien que les segments de 10 000 points ne donnaient pas les meilleurs résultats pour les données électriques, la valeur a quand même été choisie pour rester uniforme dans les tests. De plus, la troisième itération de tests pour le RF a été fait avec des segments de même longueur pour rester uniforme pour les comparaisons. L'ensemble de données utilisé est celui sur l'aluminium seulement, ce qui fait que l'apprentissage par transfert n'a pas été utilisé. Le nombre de pas a toujours été de 1, car les meilleurs modèles (acoustiques) étaient plus précis avec ce nombre de pas.

TABLEAU 5.34 : Exactitude selon la source de données

Type de données	Exactitude
Données électriques	0,511
Données acoustiques	0,645

Le Tableau 5.34 indique que les données acoustiques ont obtenu une meilleure exactitude que les données électriques. Le Tableau 5.35 montre que la classification combinée avec

TABLEAU 5.35 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,671
Classification basée sur la procédure	0,486

TABLEAU 5.36 : Exactitude selon la source de données et le type de classe

Type de données	Type de classe	Exactitude
Données électriques	Classification combinée avec diagnostic	0,649
	Classification basée sur la procédure	0,373
Données acoustiques	Classification combinée avec diagnostic	0,692
	Classification basée sur la procédure	0,598

diagnostic atteint une exactitude supérieure à la classification basée sur la procédure. Le Tableau 5.36 montre une analyse plus détaillée en combinant la source de données et le type de classe. Les données acoustiques combinées avec la classification combinée avec diagnostic affichent l'exactitude la plus élevée.

Cette troisième itération a été exécutée une deuxième fois, de manière complètement identique, avec l'ensemble de données de soudures d'acier. Il n'y a donc toujours pas eu d'apprentissage par transfert.

TABLEAU 5.37 : Exactitude selon la source de données

Type de données	Exactitude
Données électriques	0,524
Données acoustiques	0,736

TABLEAU 5.38 : Exactitude selon le type de classe

Type de classe	Exactitude
Classification combinée avec diagnostic	0,676
Classification basée sur la procédure	0,584

TABLEAU 5.39 : Exactitude selon la source de données et le type de classe

Type de données	Type de classe	Exactitude
Données électriques	Classification combinée avec diagnostic	0,609
	Classification basée sur la procédure	0,439
Données acoustiques	Classification combinée avec diagnostic	0,743
	Classification basée sur la procédure	0,729

Le Tableau 5.37 indique que les données acoustiques ont encore obtenu une meilleure exactitude que les données électriques. Le Tableau 5.38 montre que la classification combinée avec diagnostic atteint toujours une exactitude supérieure à la classification basée sur la procédure. Le Tableau 5.39 montre une analyse plus détaillée en combinant la source de données et le type de classe. Les données acoustiques associées à la classification combinée avec diagnostic ont obtenu encore l'exactitude la plus élevée. Ces résultats sont très semblables à ceux sur l'ensemble de données de l'aluminium, ce qui indique que les modèles seraient capables de s'adapter à différents ensembles de données sans modification d'architecture.

Une dernière itération a été faite dans le but de tester l'apprentissage par transfert entre les deux ensembles de données sur l'aluminium et l'acier. Pour ce faire, des modèles ont été entraînés sur un ensemble. Par la suite, toutes les couches des modèles ont été gelées, sauf un certain nombre parmi les dernières sont restées entraînables, ce qui a permis d'entraîner sur le second ensemble. Différents nombres de couches entraînables ont été testés : 4, 8, 12 et 16.

TABLEAU 5.40 : Exactitude avant et après transfert d'apprentissage pour l'aluminium vers l'acier selon le nombre de couches entraînables pour les données électriques

	Exactitude	
Couches entraînables	Pour l'aluminium	Après transfert sur l'acier
4	0,461	0,523
8	0,489	0,525
12	0,531	0,584
16	0,411	0,529

TABLEAU 5.41 : Exactitude avant et après transfert d'apprentissage pour l'acier vers l'aluminium selon le nombre de couches entraînables pour les données électriques

	Exactitude	
Couches entraînables	**Pour l'acier**	**Après transfert sur l'aluminium**
4	0,554	0,483
8	0,555	0,466
12	0,512	0,410
16	0,538	0,353

TABLEAU 5.42 : Exactitude avant et après transfert d'apprentissage pour l'aluminium vers l'acier selon le nombre de couches entraînables pour les données acoustiques

	Exactitude	
Couches entraînables	**Pour l'aluminium**	**Après transfert sur l'acier**
4	0,613	0,609
8	0,646	0,663
12	0,651	0,716
16	0,646	0,721

TABLEAU 5.43 : Exactitude avant et après transfert d'apprentissage pour l'acier vers l'aluminium selon le nombre de couches entraînables pour les données acoustiques

	Exactitude	
Couches entraînables	**Pour l'acier**	**Après transfert sur l'aluminium**
4	0,744	0,563
8	0,710	0,637
12	0,714	0,673
16	0,736	0,694

En regardant les Tableaux 5.40 et 5.42, il est facile de remarquer que les résultats semblent s'améliorer pour les données électriques et acoustiques lorsqu'on entraîne sur l'aluminium avant de transférer sur l'acier. De plus, les Tableaux 5.41 et 5.43 semblent montrer une diminution de l'exactitude lorsqu'on entraîne sur l'acier avant de faire le transfert sur l'aluminium. Cependant, l'élément vraiment important à observer dans les tableaux est de savoir si les modèles sont plus efficaces que s'ils avaient été entraînés depuis zéro. Donc, si on

compare les modèles entraînés initialement sur l'acier dans les Tableaux 5.41 et 5.43 (dans la colonne de gauche) avec les modèles entraînés sur l'aluminium avant d'être transféré sur l'acier dans les Tableaux 5.40 et 5.42 (dans la colonne de droite), il est facile de constater que l'apprentissage par transfert n'affecte pas significativement l'exactitude des modèles. L'inverse peut aussi être remarqué quand on compare les modèles entraînés sur l'aluminium avec ceux entraînés sur l'acier avant d'être transférés sur l'aluminium.

Il est également intéressant de remarquer que dans les Tableaux 5.40, 5.42, 5.41, et 5.43, la colonne de gauche représente des modèles entraînés de manière identique. Pourtant, les résultats sont un peu et même parfois beaucoup différents, ce qui indique que l'exactitude des modèles peut varier d'un entraînement à l'autre, même sans faire de modification dans la configuration. Cela signifie que, pour chaque tableau, les quatre résultats dans la colonne de gauche proviennent techniquement du même modèle, mais les exactitudes sont différentes.

5.2.1 RÉSULTATS DU CNN ADAPTÉ À DES SCALOGRAMMES

Tout comme le RF et le CNN original, le CNN adapté aux scalogrammes a été testé avec plusieurs itérations de plusieurs tests. Cependant, ces tests ont été très limités en nombre en raison de la lourdeur de calcul nécessaire pour l'entraînement d'un seul modèle qui pouvait durer plusieurs heures.

La première itération a été faite sur l'ensemble de données d'aluminium. Seule la méthode de séparation de données aléatoire a été utilisée sans état aléatoire (random state) pour réduire le temps d'exécution total. Comme il s'agissait de scalogrammes, les données étaient sous forme d'images dont les dimensions étaient toujours de 2000 par 264. Le type de classification utilisé a été celui combiné avec diagnostic avec les classes positives et négatives distinctes. Quatre types de données ont été utilisés : la tension, le courant, la combinaison

des deux pour obtenir une résistance effective instantanée, ainsi que l'amplitude acoustique. Seul la fonction d'activation ReLU a été utilisée en raison de la meilleure exactitude qu'elle a obtenu précédemment et le nombre de pas (strides) a toujours été fixé à 1. Il y a aussi le test de deux différents modes de couleurs : un en nuances de gris et l'autre en couleurs standards. Un total de 8 tests ont été effectués.

TABLEAU 5.44 : Exactitude selon le mode de couleurs avec les données d'aluminium

Mode de couleurs	Exactitude
Nuances de gris	0,700
Couleurs standards	0,668

TABLEAU 5.45 : Exactitude selon la source de données avec les données d'aluminium

Source de données	Exactitude
Courant électrique	0,645
Tension électrique	0,664
Résistance électrique	0,669
Amplitude acoustique	0,759

Le Tableau 5.44 montre que les nuances de gris ont obtenu une exactitude légèrement supérieure aux couleurs standards. Le Tableau 5.45 démontre que les données provenant de l'amplitude acoustique ont obtenu l'exactitude la plus élevée, ce qui est cohérent avec les résultats obtenues pour les modèles de CNN sur données brutes. Parmi les sources de données électriques, la résistance a montré une exactitude légèrement meilleure que les autres.

Cette première itération a été réalisée de manière identique avec l'ensemble de données de l'acier.

Le Tableau 5.46 montre que les nuances de gris ont obtenu une exactitude très légèrement supérieure à celle des couleurs standards. Le Tableau 5.47 démontre que les données provenant de l'amplitude acoustique ont encore obtenu l'exactitude la plus élevée. Pour les sources de données électriques, la résistance a encore démontré une exactitude supérieure aux autres.

TABLEAU 5.46 : Exactitude selon le mode de couleurs avec les données d'acier

Mode de couleurs	Exactitude
Nuances de gris	0,684
Couleurs standards	0,680

TABLEAU 5.47 : Exactitude selon la source de données avec les données d'acier

Source de données	Exactitude
Courant électrique	0,642
Tension électrique	0,651
Résistance électrique	0,682
Amplitude acoustique	0,753

Une deuxième et dernière itération a été effectuée avec seulement les données d'amplitude acoustique et de résistance électrique, et ce, seulement avec les images en nuances de gris. Le but était de comparer les données électriques et acoustiques avec les meilleures configuration. Cette fois-ci, des états aléatoires (random state) ont été utilisés pour effectuer chaque test trois fois. Au total, 6 tests ont été effectués sur l'ensemble de données de l'aluminium.

TABLEAU 5.48 : Exactitude selon la source de données

Source de données	Exactitude
Amplitude acoustique	0,746
Résistance électrique	0,677

Le Tableau 5.48 démontre que les données acoustiques ont obtenu une meilleure exactitude que les données électriques.

Cette itération a aussi été effectuée de manière complètement identique avec l'ensemble de données de l'acier.

Le Tableau 5.49 démontre que les données acoustiques ont obtenu une meilleure exactitude que les données électriques, même avec l'ensemble de données de l'acier.

TABLEAU 5.49 : Exactitude selon la source de données

Source de données	Exactitude
Amplitude acoustique	0,780
Résistance électrique	0,675

5.3 ANALYSE ET COMPARAISON DES RÉSULTATS

Cette section sert à analyser les résultats obtenus à l'aide des différents modèles de classification. Étant donné le grand nombre de résultats présentés dans la section précédente, seuls les plus significatifs sont discutés ici, pour éviter d'alourdir l'interprétation. L'objectif est surtout de comparer le modèle de CNN avec le RF, les données électriques avec les données acoustiques et les méthodes de classification, ainsi que de comprendre pourquoi certaines classes sont plus faciles à classifier.

5.3.1 COMPARAISON DES MÉTHODES DE CLASSIFICATION

Tel que vu dans les Tableaux 5.15, 5.19, 5.36 et 5.39, la classification combinée avec diagnostic donne de meilleures exactitudes que la classification basée sur la procédure. Cela indiquerait que les étiquettes venant de la classification combinée avec diagnostic correspondraient mieux aux motifs présents dans les signaux, car ils reflètent des défauts effectivement observés, plutôt que des variations de paramètres nominaux qui ne causent pas nécessairement des défauts dans la soudure finale. Cependant, il est possible qu'une partie de cette amélioration soit liée à une répartition déséquilibrée des classes. En effet, la classe des soudures réussies est devenue majoritaire dans la classification combinée, ce qui peut biaiser l'exactitude globale. Pour appuyer cela, le Tableau 5.50 démontre que la majorité des classes ont des F1-Score inférieurs à 50 %, malgré que l'exactitude globale soit de 69 %, ce qui indique que

ces anomalies auraient pu passer pour de bonnes soudures dans un contexte réel de production, ce qui peut représenter un enjeu important de qualité.

TABLEAU 5.50 : Rapport de classification pour un modèle de CNN sur données acoustiques avec classification combinée avec diagnostic

Classe	Précision	Rappel	F1-Score	Support
Diamètre fil d'apport	1,00	0,97	0,99	225
Type de fil d'apport	0,88	0,99	0,93	300
Hors-position +	0,40	0,51	0,45	300
Hors-position -	0,55	0,26	0,35	600
Réussite	0,72	0,87	0,78	2250
Débit de gaz +	0,43	0,37	0,40	150
Débit de gaz -	0,27	0,28	0,27	75
Longueur d'arc	0,77	0,48	0,59	450
Épaisseur	0,00	0,00	0,00	75
Exactitude (accuracy)			0,69	4425
Macro moyenne	0,56	0,52	0,53	4425
Moyenne pondérée	0,67	0,69	0,67	4425

De plus, pour une configuration équivalente (même type de données et même modèle), la différence est plus élevée lorsque la classification basée sur la procédure donne des résultats moins bons. Cela pourrait s'expliquer par le fait que cette approche inclut des anomalies procédurales qui n'affectent pas nécessairement la qualité finale des soudures. Ces soudures sont donc étiquetées comme défectueuses alors qu'elles sont en réalité conformes, ce qui introduirait de la confusion dans l'apprentissage du modèle. La classification combinée avec diagnostic corrige en partie ce problème en regroupant ces soudures non problématiques avec les soudures de référence. Cela rendrait les classes plus cohérentes du point de vue des signaux mesurés, et permettrait donc au modèle d'être plus précis. Donc, plus les performances sont faibles avec la classification basée sur la procédure, plus il serait probable que de nombreuses soudures créent de la confusion. Le gain d'exactitude observé avec la classification combinée

s'expliquerait donc par une meilleure correspondance entre les étiquettes et les caractéristiques réellement détectables dans les données.

5.3.2 COMPARAISON DES MODÈLES DE CNN ET DE RF

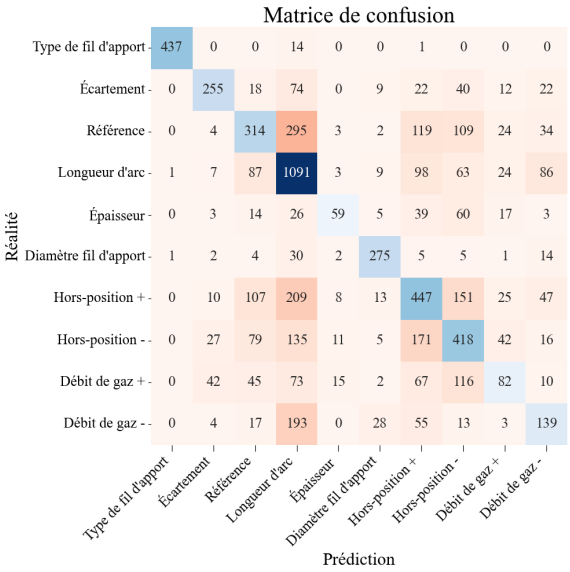
La Figure 5.1 démontre un phénomène plutôt intéressant (phénomène qui pouvait aussi être observé dans les Tableaux 5.15, 5.19, 5.36 et 5.39). Le modèle de RF obtient de meilleurs résultats lorsque les données sont électriques et le CNN a des meilleurs résultats lorsque les données sont acoustiques.

Tel que vu dans les Figures 4.2 et 4.3, les signaux électriques ont une structure plus régulière et pulsée que ceux acoustiques, tandis que les signaux acoustiques semblent moins périodiques. L'extraction de caractéristiques effectuée par TSFresh se base principalement sur des descripteurs temporels, statistiques et fréquentiels (ce qui inclut des FFT). Ces caractéristiques sont efficaces pour capturer les régularités présentes dans les signaux électriques, ce qui rend les différentes classes plus facilement séparables par des modèles comme le RF. De l'autre côté, les signaux acoustiques, de nature plus complexe et bruitée, sont plus difficiles à interpréter avec des caractéristiques classiques. Les CNN, avec leur nature « boîte noire », sont capables d'apprendre des motifs complexes et subtils à partir des données brutes qui n'auraient probablement pas pu être détectés par des méthodes traditionnelles. Cela serait donc les raisons pourquoi le CNN était meilleur avec des données acoustiques et le RF avec les données électriques.

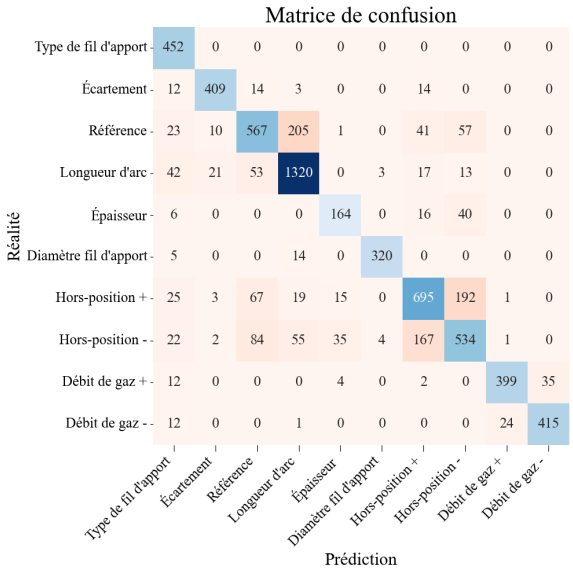
Cependant, lors de la première itération de tests pour le modèle de CNN, les signaux électriques donnaient de meilleurs résultats que les signaux acoustiques, tel que le démontre le Tableau 5.25. Cela s'expliquerait par le fait que lorsque certaines configurations étaient fixées à la valeur la plus optimale à chaque nouvelle itération, il est possible que les données électriques

étaient plus efficaces avec des configurations qui étaient, en moyenne moins bonnes, ce qui leur aurait causé une dégradation des performances en même temps qu'une augmentation pour les données acoustiques. Par exemple, le Tableau 5.32 démontre que les données acoustiques avaient de meilleures performances lorsque les données étaient de 12 000 points avec 1 pas pour la première convolution et, comme c'était le meilleur modèle, l'itération de tests suivante a été réalisée avec cette configuration. Le Tableau 5.31, de son côté, montre que les données électriques performaient mieux avec des segments de 2000 points ou des segments de 10 000 points avec 4 pas pour la première convolution. Cela indique que, si les ensembles de données avaient été optimisés de manière indépendante, les données électriques auraient pu avoir de meilleurs résultats. Par contre, cela aurait pu rendre la comparaison plus difficile en raison des configurations qui ne seraient pas uniformes entre les deux types de données.

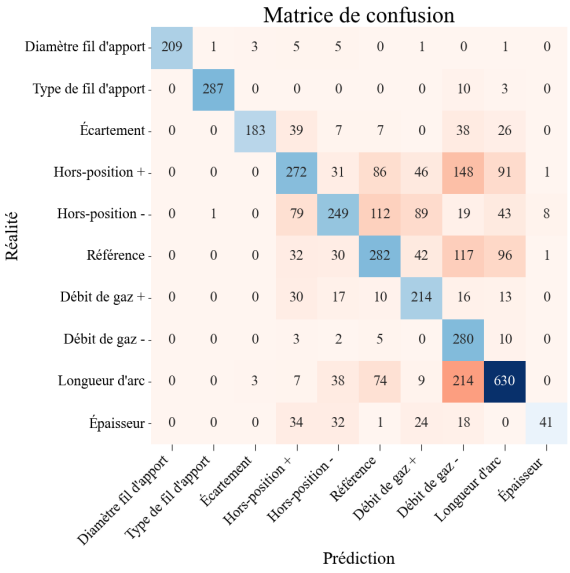
Il est aussi facile de remarquer dans la Figure 5.1 que le RF a, en moyenne, de meilleurs résultats que le CNN (66 % et 48,5 % respectivement). Il est possible que cette différence de performance provienne du fait que le nombre de soudures pour le projet était limité et que les algorithmes de DL sont généralement plus précis que les algorithmes de ML classique seulement lorsque le nombre de données est suffisamment élevée (Sarker, 2021).



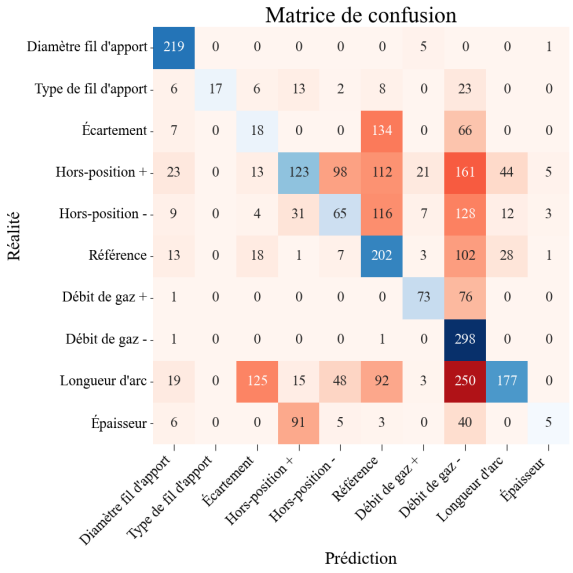
Matrice de confusion pour le modèle de RF avec les données acoustiques (exactitude de 53 %).



Matrice de confusion pour le modèle de RF avec les données électriques (exactitude de 79 %).



Matrice de confusion pour le modèle de CNN avec les données acoustiques (exactitude de 60 %).



Matrice de confusion pour le modèle de CNN avec les données électriques (exactitude de 37 %).

FIGURE 5.1 : Comparaison entre les modèles de RF et de CNN ainsi qu'entre les données acoustiques et électriques avec classification basée sur la procédure sur l'ensemble de données de l'aluminium.

5.3.3 COMPARAISON DES PERFORMANCES DES DIFFÉRENTES CLASSES

Lors de l'entraînement des modèles, les différentes classes ont eu des taux de classification différents. Pour observer cette performance, le Tableau 5.51 représente une somme des 4 matrices de confusion, qui peuvent être observées dans la Figure 5.1, a été faite pour obtenir un rapport de classification qui représente la moyenne des 4 modèles.

TABLEAU 5.51 : Rapport de classification moyen (CNN et RF avec données acoustiques et électriques.)

Classe	Précision	Rappel	F1-Score	Support
Diamètre fil d'apport	0,875	0,905	0,886	1128
Type de fil d'apport	0,906	0,935	0,896	1279
Écartement	0,738	0,602	0,659	1429
Hors-position +	0,551	0,462	0,498	3309
Hors-position -	0,518	0,452	0,476	2783
Référence	0,529	0,492	0,504	2783
Débit de gaz +	0,617	0,565	0,577	1354
Débit de gaz -	0,505	0,753	0,556	1504
Longueur d'arc	0,673	0,693	0,659	4642
Épaisseur	0,625	0,357	0,424	752

Comme le Tableau 5.51 le démontre, les classes « Diamètre fil d'apport » et « Type de fil d'apport » sont celles qui ont le mieux performé avec des F1-Score légèrement sous les 90 % et les classes « Écartement » et « Longueur d'arc » ont aussi plutôt bien performé avec des F1-Score aux alentours de 65 %. Cela pourrait s'expliquer par le fait que ces classes auraient des caractéristiques qui sont plus distinctes par rapport aux autres.

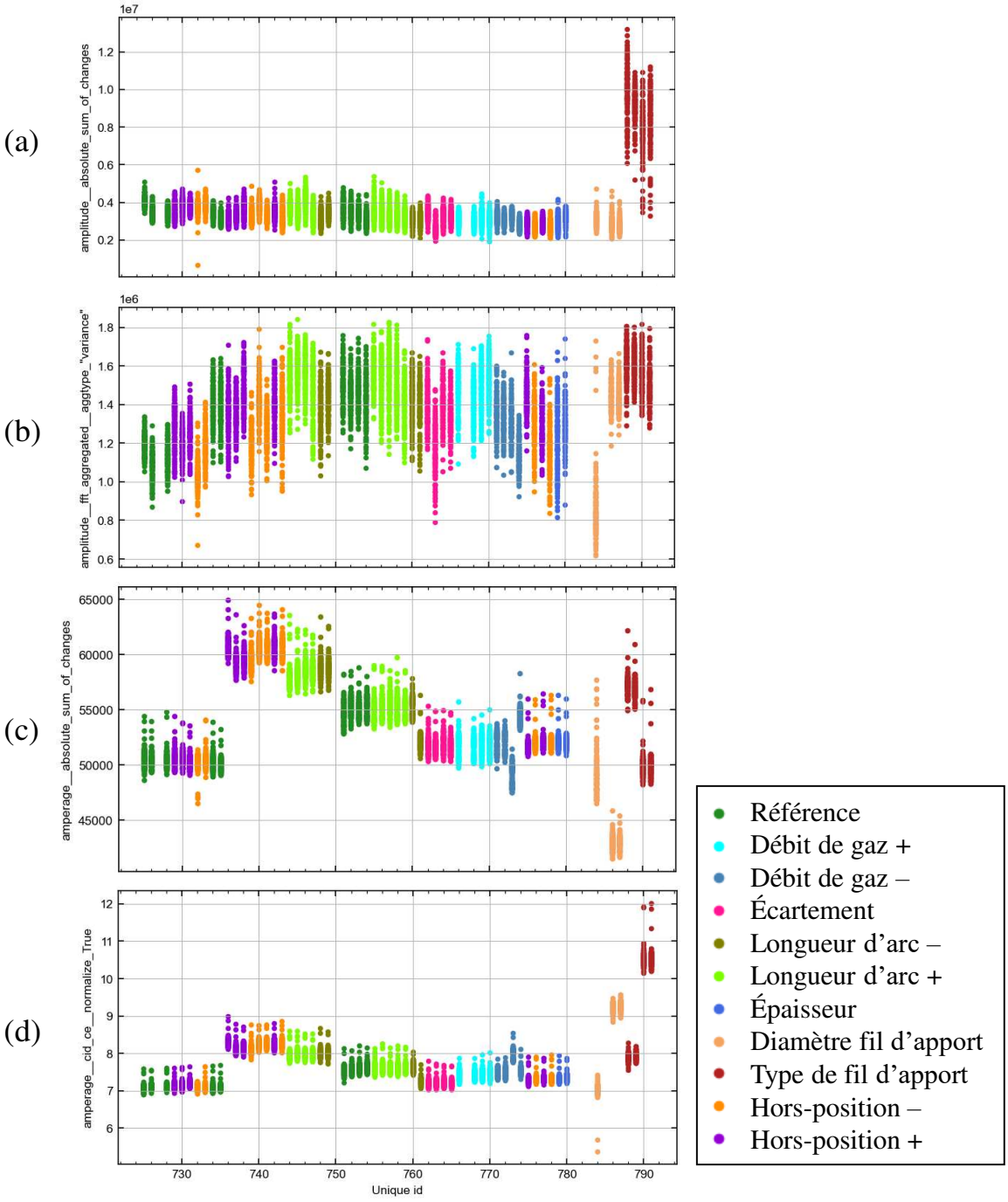


FIGURE 5.2 : Illustration des valeurs de certaines caractéristiques des données acoustiques (a-b) et électriques (c-d) extraites par TSFresh pour chaque soudure.

La Figure 5.2 démontre les valeurs de quelques caractéristiques importantes extraites par TSFresh pour des données de signaux acoustiques et électriques. L'axe des X représente un identifiant de soudure, l'axe des Y représente la valeur d'une caractéristique et chaque point représente un segment de soudure identifié par une couleur représentant une classe. Dans la droite des graphiques, il est facile de remarquer que les classes « Diamètre fil d'apport » (en orange pâle) et « Type de fil d'apport » (en rouge) ont des valeurs qui se distinguent des autres classes, ce qui pourrait expliquer pourquoi elles ont été beaucoup mieux prédites que les autres.

Il est important de remarquer que la Figure 5.2c-d montre que les valeurs extraites n'ont pas l'air similaires pour les soudures d'une même classe, mais plutôt pour les soudures qui ont été faites séquentiellement. En effet, des groupes de 6 à 11 soudures consécutives ont des valeurs semblables, sans que la classe ne semble faire de différence. Ces groupes sont en fait des soudures qui ont été faites dans une même journée, ce qui indique qu'il y a probablement du bruit qui influence les signaux électriques. Cela pourrait donc introduire un biais dans les données, ce qui nuirait aux performances des modèles et surtout à leur transférabilité.

5.3.4 EXPLICABILITÉ DES RÉSULTATS

L'explicabilité des modèles était plutôt importante, puisqu'elle peut donner confiance aux opérateurs d'utiliser les résultats pour corriger le procédé lorsqu'une anomalie est détectée. Les modèles de CNN, en raison de leur nature de DL, ne permettent pas d'obtenir une interprétation claire de leurs prédictions. Les RF, de leur côté, sont en mesure de fournir une certaine explicabilité sur les caractéristiques qui sont prises en compte dans la classification. Il est possible d'obtenir une liste des caractéristiques en ordre d'importance, et il est également possible d'obtenir un affichage visuel d'un des arbres de la forêt pour voir quelles décisions sont prises en fonction de la valeur de certaines caractéristiques.

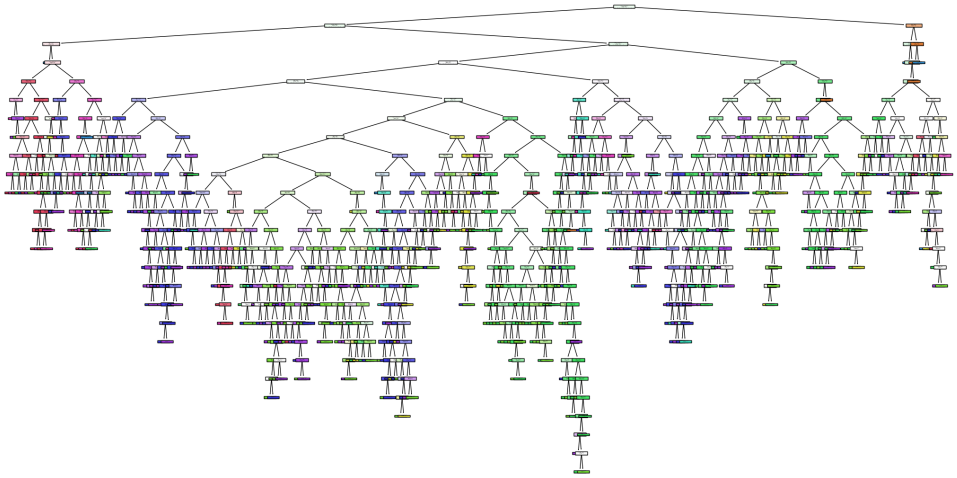


FIGURE 5.3 : Exemple d'arbre de décision venant d'un modèle de RF entraîné sur les données électriques.

Cependant, la liste d'importance des caractéristiques présentée au Tableau A.1 ne permet pas à elle seule d'expliquer l'ensemble des mécanismes internes du modèle et la Figure 5.3 montre qu'un arbre de décision peut devenir très complexe lorsque le nombre de caractéristiques est élevé, ce qui peut rendre la compréhension difficile.

Pour aider à obtenir des explications plus claires sur ce qui se passe dans la prise de décision d'un RF, la bibliothèque SHAP (Lundberg & Lee, 2017) de Python permet de mesurer à quel point une variable a influencé une prédiction. La Figure 5.4 montre les caractéristiques les plus importantes selon SHAP et elle permet aussi de voir, par code de couleur, l'importance pour les différents types d'anomalies. De plus, les Figures A.1 à A.10 montrent, pour chaque classe de la classification basée sur la procédure, l'influence des caractéristiques sur une prédiction. Dans ces figures, la longueur de chaque flèche indique l'ampleur de l'influence

d'une caractéristique à la probabilité de la classe prédite. Une flèche rouge indique une caractéristique qui augmente la probabilité de la classe et une flèche bleue indique l'inverse.

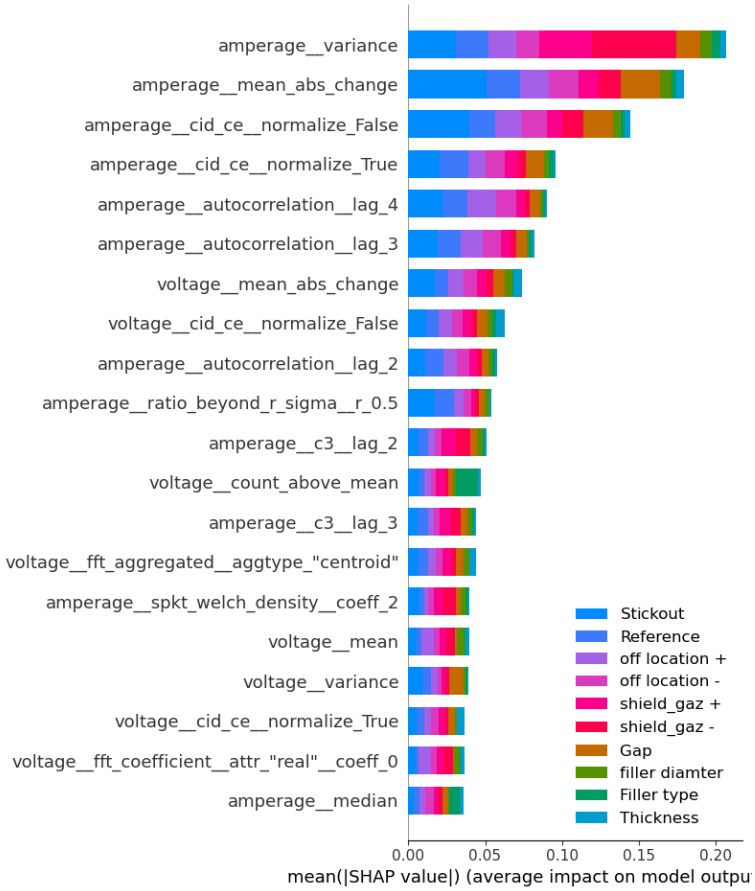


FIGURE 5.4 : Exemple d'importance globale des caractéristiques calculée par SHAP pour un RF entraîné sur les données électriques.

Compte tenu des éléments discutés précédemment, bien qu'un CNN puisse parfois être meilleur qu'un RF en performance de classification, l'explicabilité de ce dernier reste un critère important dans le choix d'un modèle à déployer en production. Dans un environnement industriel, la capacité de comprendre les facteurs qui influencent une décision est importante pour assurer une intervention rapide lorsqu'une anomalie est détectée.

Cependant, il serait aussi important de distinguer la corrélation de la causalité. En ML, certaines caractéristiques peuvent apparaître comme fortement liées à un défaut donné, sans en être la cause. En effet, la différence devrait être faite pour pouvoir faire des ajustements sur le procédé, ce qui permettrait d'éviter de faire des ajustements inefficaces ou contre-productifs. Donc, un modèle de RF qui utilise SHAP pour expliquer l'influence des caractéristiques devrait être analysé davantage pour être en mesure de confirmer que les variables les plus importantes sont réellement les causes des anomalies.

5.3.5 CONTRAINTES DE L'INDUSTRIE

Dans un contexte industriel, le déploiement d'un système de détection d'anomalies a plusieurs contraintes importantes à prendre en considération. Un modèle qui a une excellente exactitude en laboratoire peut tout de même être inadapté pour une mise en production si le nombre de données nécessaire à son apprentissage, le coût de calcul pour atteindre un niveau d'exactitude élevé ou les besoins en maintenance et en formation des opérateurs sont supérieurs aux potentiels bénéfices économiques que le modèle peut amener. De plus, pour remplacer les méthodes de détection actuelles, le modèle doit atteindre un seuil d'exactitude suffisamment élevé pour éviter qu'un trop grand nombre d'anomalies passent en production, ce qui aurait un impact économique direct (Blondheim, 2022). Selon la nature d'une anomalie et son influence sur la qualité d'une soudure, un certain taux de faux positifs ou de faux négatifs peut être toléré, mais si l'exactitude est insuffisante, le système ne serait pas économiquement viable.

Un autre aspect plutôt important est la fréquence d'échantillonnage utilisée pour l'acquisition des signaux. Une fréquence élevée, comme 40 kHz, permet de capturer des détails précis du signal, mais génère un volume de données important, ce qui amène des coûts plus élevés en stockage, en transmission et en traitement. Dans certains cas, un taux d'échantillonnage

plus bas pourrait être suffisant pour maintenir des performances acceptables, tout en réduisant significativement les coûts et la complexité du système, ainsi que les coûts des capteurs. Les modèles comme le RF, qui exploitent des caractéristiques préalablement extraites, pourraient réagir différemment à une réduction du taux d'échantillonnage que les CNN qui travaillent avec des signaux bruts. Il pourrait donc être intéressant d'évaluer l'impact de différents taux d'échantillonnage sur ces deux types de modèles, afin de déterminer si l'un est plus tolérant que l'autre dans le contexte de détection d'anomalies de soudures. Cela pourrait donc aider à développer un système qui serait économiquement viable dans l'industrie.

5.4 GÉNÉRALISATION APRÈS CHANGEMENTS DANS L'INSTALLATION

L'un des objectifs du projet était d'avoir une certaine capacité à généraliser, c'est-à-dire de maintenir de bonnes performances lorsque les conditions sont légèrement différentes de celles rencontrées durant l'entraînement des modèles. Afin d'évaluer cela, des tests ont été réalisés à partir de soudures produites dans un environnement légèrement différent : le robot de soudage a été remplacé par un modèle différent (Fanuc M700iC aussi de 45 kg de capacité), et l'emplacement physique du banc d'essai a été changé. La source de soudage n'a pas été modifiée, ce qui signifie que, en théorie, le procédé serait identique à ce qu'il était précédemment. Les systèmes de collecte de données étaient aussi les mêmes, mais ont été relocalisés. Il était attendu que les modèles maintiennent des capacités prédictives similaires sur les deux montages.

Cependant, les modèles entraînés uniquement sur les anciennes données (installation initiale) n'ont pas réussi à obtenir de bonnes prédictions sur les nouvelles soudures. Chaque modèle prédisait toujours des classes incorrectes, peu importe la vraie anomalie de la soudure. Le comportement était légèrement différent d'un modèle à l'autre, mais un même modèle donnait toujours le même genre de prédiction erronée (les classes prédites étaient pratiquement

les mêmes d'une soudure à l'autre). Ce comportement semble indiquer une incapacité des modèles à différencier les classes avec la nouvelle installation, ce qui est probablement liée à un changement dans les caractéristiques des données en raison de la nouvelle configuration.

Il est probable que ce changement dans les données soit dû à des variations dans le procédé, telles que le positionnement du robot et la sensibilité des capteurs, ainsi qu'à des interférences, comme le bruit ambiant et les vibrations mécaniques, qui ont un impact sur les signaux électriques et acoustiques des soudures. Cela montre que des ajustements qui pourraient paraître sans incidence sur le procédé peuvent modifier la structure des signaux mesurés, et ce, suffisamment pour nuire aux performances des modèles.

Pour tenter de trouver une solution, un nouveau modèle de RF a été entraîné à partir d'un petit sous-ensemble de données venant de la nouvelle installation (environ 5 ou 6 soudures), cette fois avec une classification binaire simplifiée (avec les classes Référence et Longueur d'arc). Le RF a été choisi parce qu'il a le potentiel d'obtenir des résultats relativement bons avec peu de données, contrairement au CNN qui a généralement besoin de plus de données pour être performant. Malgré la taille restreinte du jeu de données, les prédictions sur deux ou trois soudures supplémentaires issues de cette même configuration ont été jugées satisfaisantes. Aucune métrique quantitative n'a été calculée en raison du nombre de données trop faible, mais qualitativement, les segments de test semblaient être bien classés.

Ces résultats mettent de l'avant deux points importants :

— Les modèles provenant de la configuration initiale ne sont pas capables de bien généraliser lorsqu'il y a des variations mineures dans la configuration expérimentale. En d'autres mots, le même modèle pourrait ne pas être applicable sur deux cellules nominalement identiques adjacentes, ce qui limite leur déploiement dans des environnements de production réels.

— Un réentraînement sur un petit échantillon de nouvelles données peut toutefois permettre d'obtenir une performance acceptable pour un objectif simplifié de classification binaire, notamment pour des modèles comme un RF qui sont relativement rapides à entraîner par rapport à des modèles de DL.

Ces résultats montrent l'importance de développer des modèles capables de généraliser à partir de données provenant de différentes installations à partir d'un minimum de données d'entraînement, ce qui permet ainsi de créer un modèle applicable dans plusieurs contextes industriels différents.

CHAPITRE VI
APPLICATION À LA CELLULE DE SOUDAGE

Bien que les modèles soient capables de détecter des anomalies, il est surtout important de pouvoir intégrer les modèles dans un environnement de production pour que des opérateurs puissent intervenir si nécessaire. Afin de démontrer le potentiel d'intégration industrielle des modèles, une application visuelle a été développée à l'aide de la bibliothèque Streamlit de Python, ce qui permet d'obtenir un affichage simple à partir d'une interface web.

Cette application est directement connectée à la base de données des soudures. Elle est conçue pour s'activer automatiquement dès qu'une nouvelle soudure est complétée. Lorsque cela est détecté, l'application extrait les données associées à cette soudure et exécute les prédictions à l'aide du modèle de classification sélectionné.

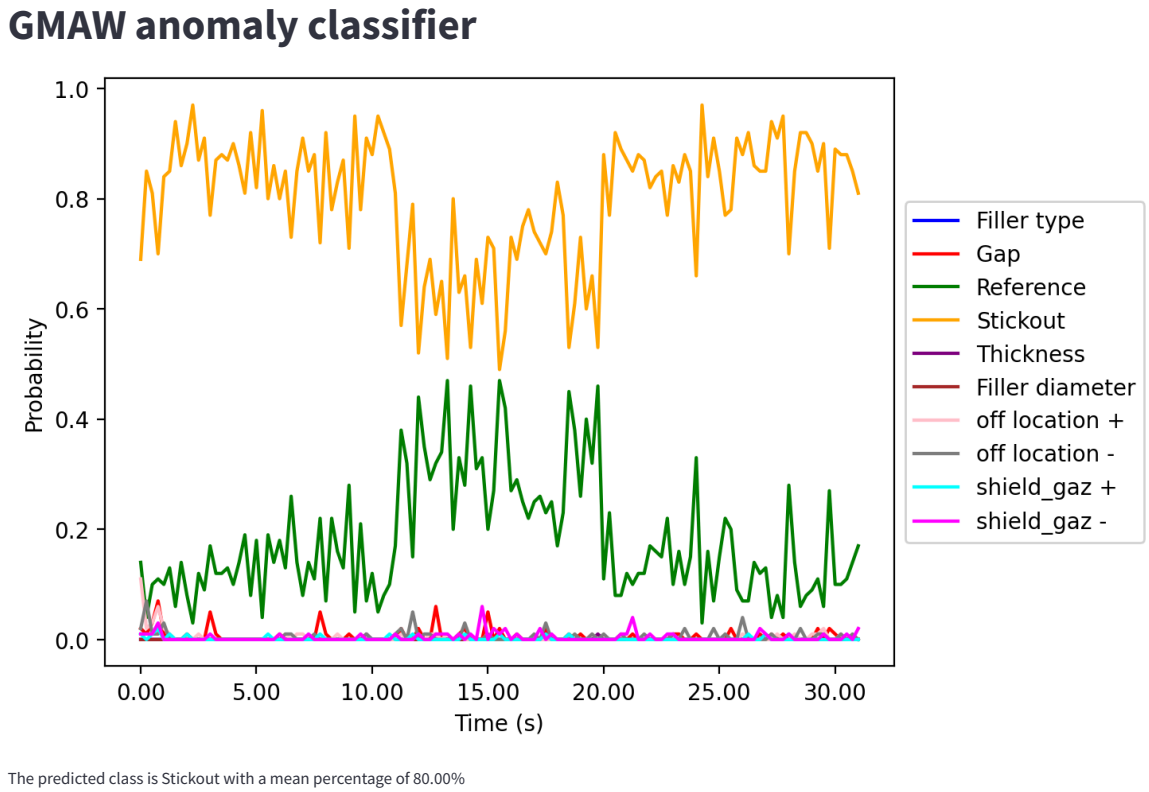


FIGURE 6.1 : Exemple de classification correcte d'une soudure avec anomalie de longueur d'arc. Classification effectuée avec un RF sur données électriques de 10 000 points.

Comme la Figure 6.1 le montre, les résultats sont présentés sous forme graphique, où chaque segment de la soudure est représenté par un pourcentage de prédiction pour chacune des classes possibles. Cela permet d'avoir une visualisation détaillée des prédictions le long de la soudure, ce qui facilite l'identification des segments qui contiennent une anomalie. De plus, une moyenne pondérée des pourcentages de prédiction de tous les segments est calculée et affichée en bas de la figure afin de générer une prédiction globale pour la soudure complète. Cela donne un aperçu des probabilités associées à chaque classe, ce qui permet aux opérateurs de mieux comprendre la confiance du modèle dans ses prédictions et d'ajuster les décisions en conséquence. Étant donné que les prédictions sont disponibles pour chaque segment, une approche orientée sur les défauts localisés serait aussi réalisable.

Toujours dans la Figure 6.1, il est facilement remarquable que les données du milieu de la soudure semblent moins bien prédites. Cela est dû au fait que le modèle avait été entraîné avec les segments du milieu comme données de test avec la méthode de séparation par sections de soudures (voir Figure 4.4), ce qui fait que les données du milieu sont des vraies données de tests et les autres sont des données qui étaient dans l'entraînement.

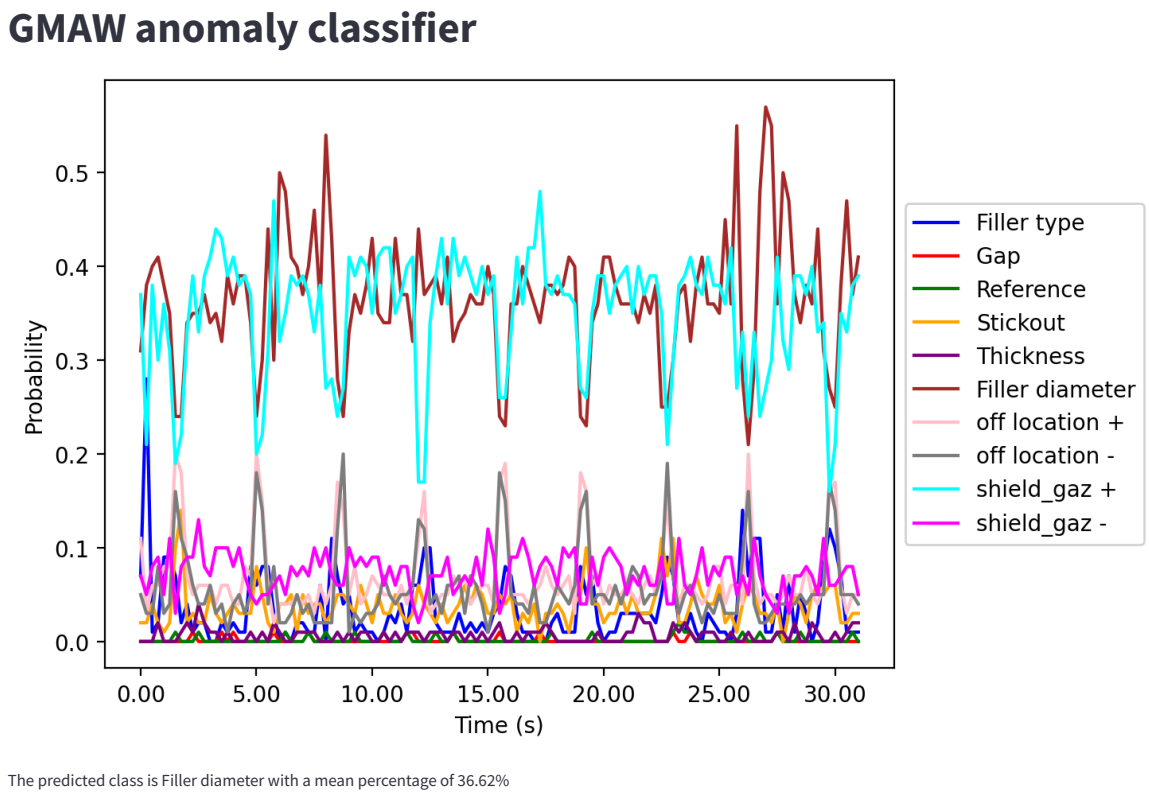


FIGURE 6.2 : Exemple de classification incorrecte d'une soudure de référence effectuée après changements dans l'installation.
Classification effectuée avec un RF sur données électriques de 10 000 points.

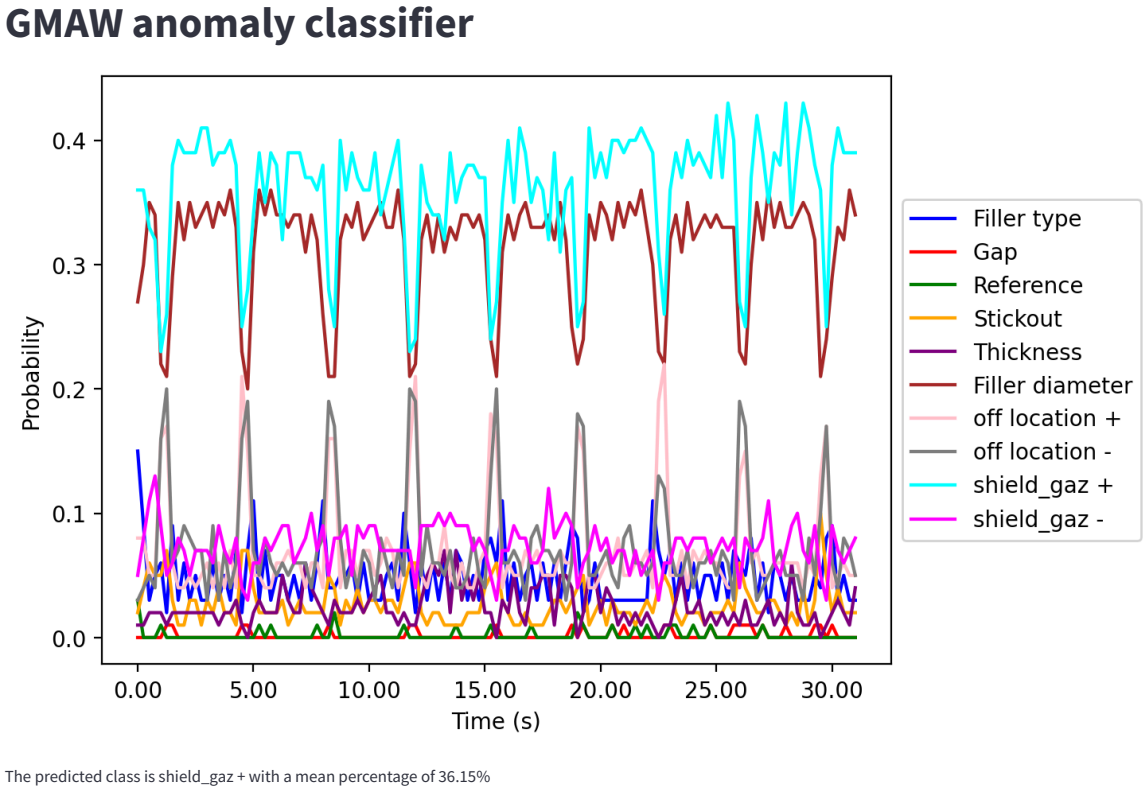


FIGURE 6.3 : Exemple de classification incorrecte d'une soudure avec anomalie de longueur d'arc effectuée après changements dans l'installation.
Classification effectuée avec un RF sur données électriques de 10 000 points.

Tel que mentionné dans la section 5.4, les classifications des modèles étaient mal effectuées avec de nouvelles soudures qui ont été faites après certains changement dans l'installation de soudage. La soudure rapportée à la Figure 6.2 aurait dû se faire classifier comme étant une référence et la Figure 6.3 aurait dû afficher une anomalie de longueur d'arc, mais les deux ont des prédictions incorrectes qui se ressemblent visuellement. Ces deux figures montrent donc visuellement les problèmes de généralisation après les changements dans l'installation.

Il est intéressant de remarquer que dans les Figures 6.2 et 6.3, il y a des pics constants à chaque intervalle d'environ 3 secondes où les prédictions changent. Comme le montre

également la Figure 6.4, ces pics coïncident avec des variations similaires dans le signal de courant (les Figures 6.3 et 6.4 représentent la même soudure). La comparaison avec la Figure 4.2 montre qu'aucun de ces pics n'est présent avant le changement d'installation, ce qui semble confirmer que les signaux électriques ont été modifiés par ce changement. Ces observations illustrent l'importance capitale de la qualité et fiabilité des données d'entrées aux modèles d'intelligence artificielle développés ici.

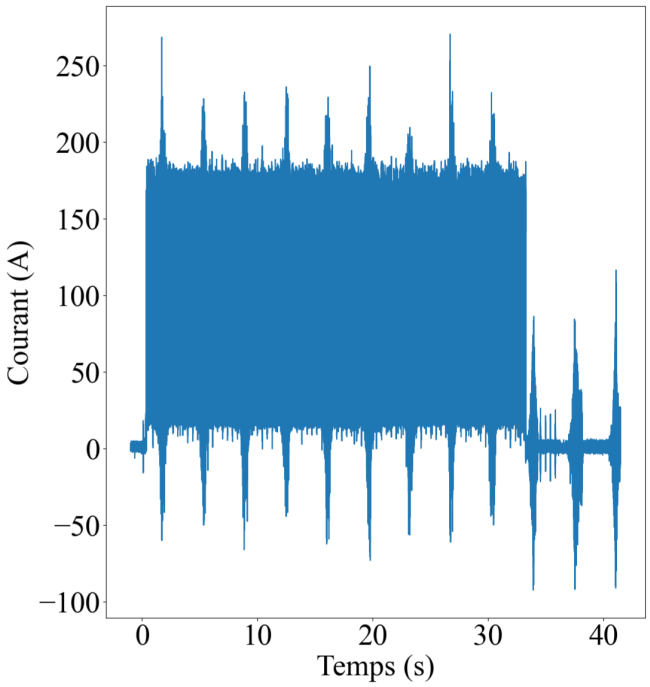


FIGURE 6.4 : Exemple de signaux de courant pour une soudure effectuée après changements dans l'installation.

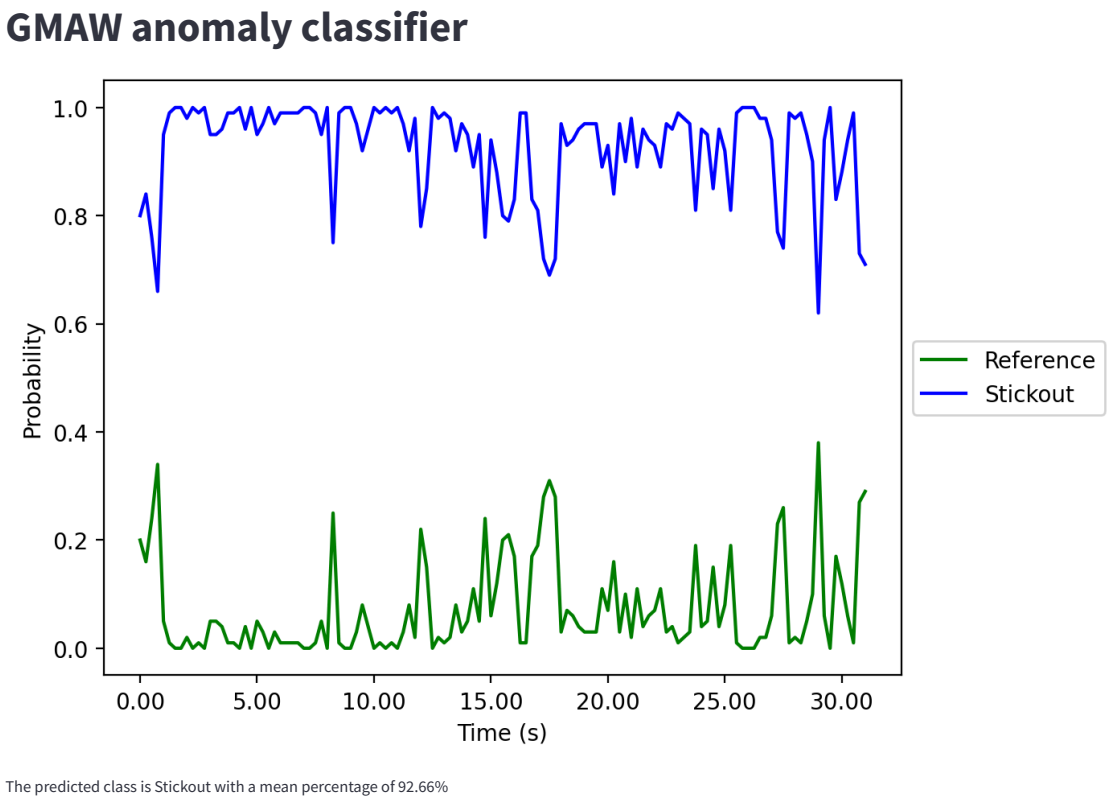


FIGURE 6.5 : Exemple de classification correcte d'une soudure avec anomalie de longueur d'arc effectuée sur un modèle entraîné après changements dans l'installation. Classification effectuée avec un RF sur données électriques de 10 000 points.

La section 5.4 avait aussi mentionné qu'un modèle entraîné avec de nouvelles soudures était capable d'obtenir des prédictions correctes pour un objectif de classification binaire sur la longueur d'arc. La Figure 6.5 montre une prédiction correcte pour une soudure de test avec une anomalie de longueur d'arc. La soudure a volontairement été gardée à l'écart pour qu'elle serve de test à 100 %, plutôt que seulement les données du milieu comme dans la Figure 6.1.

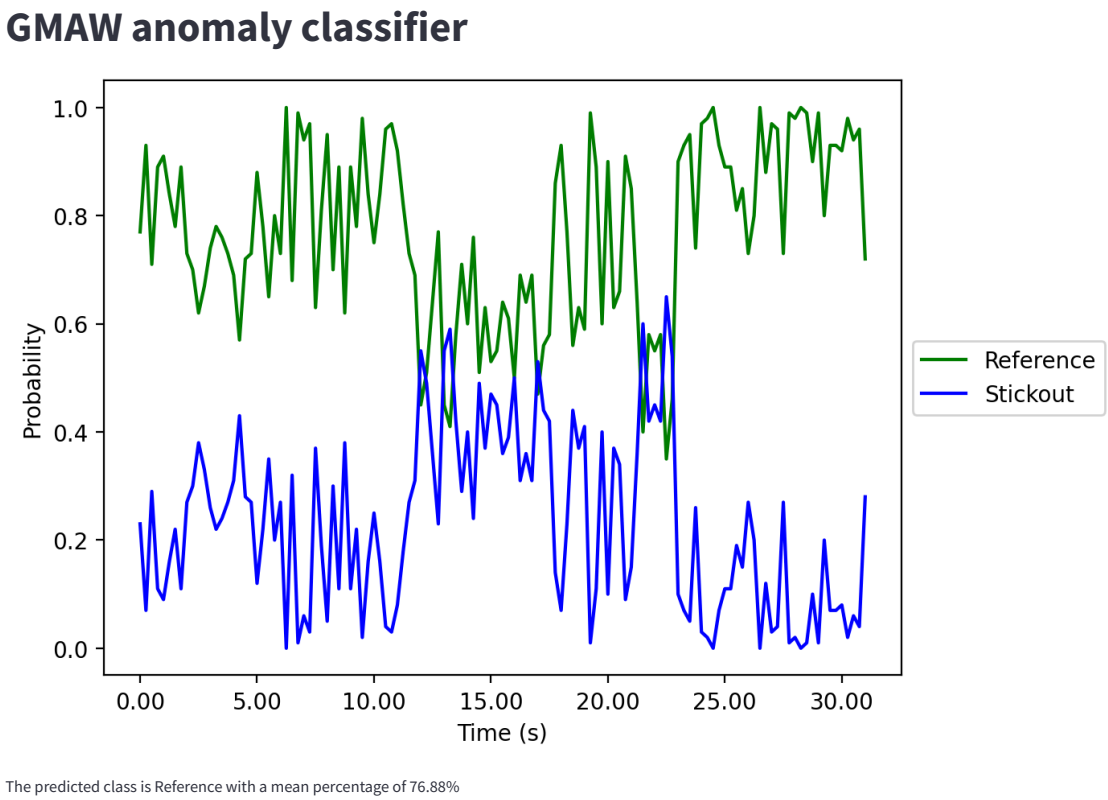


FIGURE 6.6 : Exemple de classification d'une soudure avec longueur d'arc modifié en cours de route sur un modèle entraîné après changements dans l'installation. Classification effectuée avec un RF sur données électriques de 10 000 points.

La Figure 6.6 montre une expérimentation qui a été faite avec un modèle entraîné sur les nouvelles soudures. L'objectif était de voir si le modèle était capable de détecter une anomalie transitoire qui apparaît en cours de route. La longueur d'arc a donc été légèrement modifiée au milieu de la soudure en altérant le chemin du robot et le modèle semble avoir des prédictions différentes pour cette section. Cela semble donc montrer que les modèles pourraient être capable de détecter des anomalies dans des endroits localisés dans des soudures.

Grâce à cette interface visuelle, les utilisateurs peuvent accéder facilement aux résultats de classification pour chaque soudure, ce qui pourrait aider à avoir une prise de décision rapide dans un environnement de production. En plus de la classification des soudures, l'application

permet de visualiser la variation des probabilités de chaque classe, ce qui permet aux opérateurs d'anticiper les changements potentiels du procédé et d'ajuster rapidement les paramètres en conséquence. Cependant, en raison des délais de la base de données et du temps d'extraction des caractéristiques (pour le modèle de RF seulement), le système ne permet pas d'avoir les résultats de classification en moins de 20 secondes, tel que mentionné dans la section 1.2. Malgré cela, l'intégration de cette application dans l'environnement de la cellule de soudage serait une première étape vers l'implémentation d'un système de surveillance intelligent dans des productions industrielles.

CHAPITRE VII
CONCLUSION

Le projet s'inscrit dans le contexte de l'amélioration du contrôle qualité en soudage GMAW automatisé de l'aluminium. Actuellement, les inspections reposent principalement sur des méthodes destructives ou non destructives en fin de production, telles que les analyses aux rayons X ou la métallographie optique, qui ont un certain coût et ne permettent pas d'intervenir rapidement sur le procédé. L'objectif était donc de mettre en place un système capable de détecter en quasi-temps réel des anomalies typiques du soudage, à partir de signaux électriques (courant, tension) et acoustiques qui sont collectés directement durant l'exécution.

L'objectif général du projet était de concevoir et d'implanter un système informatique de surveillance du procédé de soudage GMAW et d'en démontrer la faisabilité dans un environnement représentatif de la production industrielle. Les objectifs spécifiques étaient :

— Acquérir et prétraiter des données électriques et acoustiques associées à différents types d'anomalies ;

— Entraîner et évaluer plusieurs modèles de ML pour la classification ;

— Comparer leurs performances et leur adaptabilité aux contraintes industrielles.

Ce travail apporte plusieurs nouveautés par rapport à la littérature existante. Il propose premièrement une comparaison directe des performances des signaux électriques et acoustiques pour la détection d'anomalies en soudage. Il met également en œuvre une classification multi-classes couvrant plusieurs types d'anomalies, ce qui n'était pas fréquent dans les publications scientifiques qui faisaient souvent de la classification avec un petit nombre de classes. Il amène aussi une comparaison entre un algorithme de ML classique (RF), et une approche de DL (CNN), afin de déterminer leurs avantages respectifs dans ce contexte.

Le projet a été mené à partir d'un jeu de données collecté en environnement expérimental contrôlé, afin de garantir un étiquetage fiable des anomalies. 60 soudures GMAW ont été réalisées, dont certaines comportaient volontairement une anomalie typique. Pour chaque soudure, les signaux électriques (courant, tension) et acoustiques ont été enregistrés à haute fréquence, puis prétraités par filtrage et segmentation temporelle. Des caractéristiques ont été extraites à l'aide de TSFresh pour le RF et des signaux bruts ont été fournis directement au CNN.

Les principaux résultats montrent que les performances varient selon la combinaison du modèle du type de données. Les signaux électriques ont donné de meilleurs résultats avec le RF et ont obtenu les meilleures exactitudes globales parmi les configurations testées. Les signaux acoustiques ont été plus efficaces avec le CNN, qui a surpassé le RF sur ce type de données. Les résultats sont moins probants que certains de la littérature, mais la complexité du problème de classification multiclasse choisi pour ce projet semble plus grande. L'utilisation de SHAP a permis d'identifier les caractéristiques les plus influentes dans les décisions du RF, ce qui offre un niveau d'explicabilité difficile à obtenir avec un CNN et que les opérateurs pourraient exploiter pour corriger le procédé.

Une application visuelle, qui affiche les prédictions de chaque segment ainsi qu'une prédiction globale, a permis de démontrer le potentiel d'intégration industrielle des modèles. Les essais ont montré certaines limites de généralisation après modifications de l'installation, mais aussi un modèle réentraîné à partir d'un ensemble de données limité a été en mesure de détecter des instances d'anomalie dans un cas simplifié de classification binaire.

7.1 PERSPECTIVES DE RECHERCHE

Plusieurs pistes de recherche non explorées ont été identifiées lors de la réalisation de ce projet dans le but d'en améliorer l'application industrielle.

Une première idée serait de développer un système qui rend les données disponibles au modèle en même temps qu'une soudure est effectuée, plutôt que de les collecter à la fin. En obtenant les segments du début d'une soudure, la détection d'anomalies pourrait être faite en temps réel, et ce, avant même que la soudure soit terminée pour permettre une rétroaction rapide.

Une autre piste d'amélioration, serait de combiner les données électriques et acoustiques pour l'entraînement d'un seul modèle, ce qui permettrait d'avoir plus d'informations pour les prédictions des modèles. Cette piste aurait théoriquement pu être explorée lors de ce projet, mais cela aurait demandé un certain travail supplémentaire en raison des données qui n'ont pas la même fréquence d'échantillonnage (40 kHz pour électrique et 48 kHz pour acoustique) et aussi, à cause de la nécessité d'un alignement temporel précis des deux sources de données.

Il serait aussi intéressant de développer des sous-modèles spécialisés dans la détection de certains types d'anomalies plus difficiles à identifier dans le cas multi-classe, plutôt qu'un seul modèle général. Cela pourrait améliorer la classification pour chaque défaut et même faciliter l'adaptation à de nouvelles classes d'anomalies.

BIBLIOGRAPHIE

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X. (2016). *TensorFlow : A system for large-scale machine learning* [Conference paper]. doi: 10.48550/arXiv.1605.08695

Abdi, H. & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews : Computational Statistics*, *2*(4), 433 – 459. doi: 10.1002/wics.101

American Welding Society (2025, juin). *What Is GMAW ? A Comprehensive Overview of the GMAW Process*. Welding Digest, June 2025. Magazine technique professionnel, Repéré à https://www.aws.org/magazines-and-media/welding-digest/wd-june-2025-what-is-gmaw/

Aminzadeh, A. (2024, avril). *Amélioration de la qualité du soudage au laser des alliages d'aluminium par une surveillance en temps réel des défauts pour un processus de soudage intelligent*. Rimouski, Québec, Canada. Thèse de doctorat, Repéré à https://semaphore.uqar.ca/id/eprint/3068/

Blondheim, D. (2022). Improving Manufacturing Applications of Machine Learning by Understanding Defect Classification and the Critical Error Threshold. *International Journal of Metalcasting*, *16*(2), 502 – 520. Cited by : 21 ; All Open Access, Green Open Access, Hybrid Gold Open Access, doi: 10.1007/s40962-021-00637-0

Bouslah, B., Gharbi, A. & Pellerin, R. (2016). Integrated production, sampling quality control and maintenance of deteriorating production systems with AOQL constraint. *Omega*, *61*, 110–126. doi: https://doi.org/10.1016/j.omega.2015.07.012

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Briand, O., Mirakhorli, F. & Gariépy, A. (2024). Cyber physical production system for durable assemblies of recycled aluminium alloys in EV structures – Welding anomaly detection case studies. Confidential report to METALTec participants.

Christ, M., Braun, N., Neuffer, J. & Kempa-Liehr, A. W. (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh - A Python package). *Neurocomputing*, *307*, 72–77. doi: https://doi.org/10.1016/j.neucom.2018.03.067

Cullen, M. & Ji, J. C. (2025). Online defect detection and penetration estimation system for gas metal arc welding. *The International Journal of Advanced Manufacturing Technology*, *136*(5), 2143–2164. doi: 10.1007/s00170-024-14932-7

Dubey, S. R., Singh, S. K. & Chaudhuri, B. B. (2022). Activation functions in deep learning : A comprehensive survey and benchmark. *Neurocomputing*, *503*, 92–108. doi: https://doi.org/10.1016/j.neucom.2022.06.111

Duongthipthewa, O., Meesublak, K., Takahashi, A. & Mitsantisuk, C. (2023). Detection Welding Performance of Industrial Robot Using Machine Learning. Dans *2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)*, pp. 1–6. doi: 10.1109/ITC-CSCC58803.2023.10212676

Elía, I. & Pagola, M. (2025). Anomaly detection in Smart-manufacturing era : A review. *Engineering Applications of Artificial Intelligence*, *139*, 109578. doi: https://doi.org/10.1016/j.engappai.2024.109578

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L. & Muller, P.-A. (2019). Deep learning for time series classification : a review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. doi: 10.1007/s10618-019-00619-1

Fondation canadienne pour l'innovation (2025). *Centre des technologies de l'aluminium - Conseil national de recherches Canada*. Navigateur d'installations de recherche. Repéré à https://navigator.innovation.ca/fr/facility/conseil-national-de-recherches-canada/centre-des-technologies-de-laluminium

Hadžihafizović, D. (2023). *Welding Automation - Robotic Welding*. SSRN Preprint Repository. Available at SSRN : https://ssrn.com/abstract=4778793

Hong, Y., Yang, M., Jiang, Y., Du, D. & Chang, B. (2023). Real-Time Quality Monitoring of Ultrathin Sheets Edge Welding Based on Microvision Sensing and SOCIFS-SVM. *IEEE Transactions on Industrial Informatics*, *19*(4), 5506–5516. doi: 10.1109/TII.2022.3199258

Huang, L., Zhang, S., Li, R., Cai, X., Zhang, S. & Cao, H. (2023). Weld Defect Detection based on Improved Multi-Scale CNN with CBAM Attention. Dans *Proceedings of the 2023 6th International Conference on Signal Processing and Machine Learning*, SPML '23, pp. 214–220., New York, NY, USA. Association for Computing Machinery. doi: 10.1145/3614008.3614042

Jin, C., Shin, S., Yu, J. & Rhee, S. (2020). Prediction Model for Back-Bead Monitoring During Gas Metal Arc Welding Using Supervised Deep Learning. *IEEE Access*, *8*, 224044–224058. doi: 10.1109/ACCESS.2020.3041274

LeCun, Y. & Bengio, Y. (1998). Convolutional Networks for Images, Speech, and Time Series. Dans *The Handbook of Brain Theory and Neural Networks* pp. 255–258. MIT Press

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi: 10.1038/nature14539

Liang, X., Wang, K. & Duan, R. (2021). Characteristic Analysis of Arc Acoustic Signal in Pipeline P-GMAW Automatic Welding. Dans *2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pp. 541–544. doi: 10.1109/AUTEEE52864.2021.9668328

Liu, H. & Setiono, R. (1995). Chi2 : feature selection and discretization of numeric attributes. Dans *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, pp. 388–391. doi: 10.1109/TAI.1995.479783

Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D. & Jones, N. S. (2019). catch22 : CAnonical Time-series CHaracteristics. *Data Mining and Knowledge Discovery*, *33*(6), 1821–1852. doi: 10.1007/s10618-019-00647-x

Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Dans *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 4768–4777., Red Hook, NY, USA. Curran Associates Inc.

Mattera, G. & Nele, L. (2025). Machine learning approaches for real-time process anomaly detection in wire arc additive manufacturing. *The International Journal of Advanced Manufacturing Technology*, pp. 2863–2888. doi: 10.1007/s00170-025-15327-y

Pattanayak, S. & Sahoo, S. K. (2021). Gas metal arc welding based additive manufacturing—a review. *CIRP Journal of Manufacturing Science and Technology*, *33*, 398–442. doi: https://doi.org/10.1016/j.cirpj.2021.04.010

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. (2011). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rohe, M., Stoll, B. N., Hildebrand, J., Reimann, J. & Bergmann, J. P. (2021). Detecting Process Anomalies in the GMAW Process by Acoustic Sensing with a Convolutional Neural Network (CNN) for Classification. *Journal of Manufacturing and Materials Processing*, *5*(4). doi: 10.3390/jmmp5040135

Sarker, I. H. (2021). Deep Learning : A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. © The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2021, doi: 10.1007/s42979-021-00815-1

Schmidt, A., Kotschote, C. & Riedel, O. (2021). Supervised learning based observer for in-process tool offset estimation in robotic arc welding applications. Dans *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pp. 1991–1996. doi: 10.1109/CASE49439.2021.9551533

Scholz, J., Holtkemper, M., Graß, A. & Beecks, C. (2024). Anomaly Detection in Manufacturing. Dans J. Soldatos (Éd.). *Artificial Intelligence in Manufacturing : Enabling Intelligent, Flexible and Cost-Effective Production Through AI* (pp. 351–360). Cham: Springer Nature Switzerland. doi: 10.1007/978-3-031-46452-2_20

Skar, E. M., Kloumann, J.-E., Robbersmyr, K. G. & Løvåsen, T. (2023). Anomaly Detection in Robotic Welds - Investigation of LSTM Autoencoder Model Performance. Dans *2023 11th International Conference on Control, Mechatronics and Automation (ICCMA)*, pp. 265–270. doi: 10.1109/ICCMA59762.2023.10374923

Thekkuden, D. T., Mourad, A.-H. I. & Sherif, M. M. (2020). Response surface analysis of statistical features of voltage and current in a GMAW powersource on welding v-groove joints. Dans *2020 Advances in Science and Engineering Technology International*

Conferences (ASET), pp. 1–6. doi: 10.1109/ASET48392.2020.9118210

Vander Voort, G. F. (2011, 6). Metallography of Welds. *Advanced Materials and Processes*, pp. 19–23.

Viviani, A., Nogueira, R., Cirino, L., Silva, R. & Sousa, G. (2019, 01). WELDING POWER SOURCES WAVEFORMS ANALYSIS. doi: 10.26678/ABCM.COBEM2019.COB2019-2201

Wang, X., Zscherpel, U., Tripicchio, P., D'Avella, S., Zhang, B., Wu, J., Liang, Z., Zhou, S. & Yu, X. (2025). A comprehensive review of welding defect recognition from X-ray images. *Journal of Manufacturing Processes*, *140*, 161–180. doi: https://doi.org/10.1016/j.jmapro.2025.02.039

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R. X., Kurfess, T. & Guzzo, J. A. (2017). A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. *Journal of Manufacturing Systems*, *43*, 25–34. doi: https://doi.org/10.1016/j.jmsy.2017.02.011

Wuest, T., Weimer, D., Irgens, C. & Thoben, K.-D. (2016). Machine learning in manufacturing : advantages, challenges, and applications. *Production & Manufacturing Research*, *4*(1), 23–45. doi: 10.1080/21693277.2016.1192517

Zheng, H., Qi, B., Yang, M. & Liu, H. (2022). Effect of Arc Behaviour and Metal Transfer Process on Aluminium Welds during Ultrasonic Frequency Pulsed GMAW. *Crystals*, *12*(5). doi: 10.3390/cryst12050586

Zheng, Y., Liu, Q., Chen, E., Ge, Y. & Zhao, J. L. (2014). Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. Dans *Web-Age Information Management*, pp. 298–310., Cham. Springer International Publishing.

APPENDICE A

DÉTAILS SUPPLÉMENTAIRES SUR L'EXPLICABILITÉ DES RÉSULTATS

TABLEAU A.1 : Exemple d'importance des caractéristiques d'un RF entraîné sur des données électriques.

Caractéristique	Importance
amperage__mean_abs_change	0.04411804144486395
amperage__autocorrelation__lag_4	0.04190681093564115
amperage__variance	0.03993323587820787
amperage__cid_ce__normalize_False	0.036199412072712496
amperage__autocorrelation__lag_3	0.03540319829547465
amperage__cid_ce__normalize_True	0.03514043009718676
amperage__autocorrelation__lag_2	0.022216593268630708
voltage__count_above_mean	0.021777713954277536
voltage__mean_abs_change	0.02046965718351105
voltage__variance	0.019877079542570085
voltage__root_mean_square	0.019195439600382284
voltage__fft_coefficient__attr_real__coeff_0	0.018633570463475243
amperage__median	0.018560916268525938
amperage__c3__lag_2	0.018316337862605815
voltage__cid_ce__normalize_False	0.018017047669408245
voltage__mean	0.017573120337436535
amperage__c3__lag_3	0.017405949933211304
voltage__c3__lag_3	0.017402901480651854

Suite à la page suivante

Tableau A.1 – suite

Caractéristique	Importance
amperage__ratio_beyond_r_sigma__r_0.5	0.017064257296292927
voltage__median	0.016717520609856002
voltage__c3__lag_2	0.0162199186402367
amperage__spkt_welch_density__coeff_5	0.015159239094983073
amperage__mean	0.01478190355791674
amperage__spkt_welch_density__coeff_2	0.014251971831860584
voltage__ratio_beyond_r_sigma__r_2.5	0.013842331183776074
voltage__fft_aggregated__aggtype_variance	0.013801369393145479
amperage__root_mean_square	0.01378785753454252
amperage__kurtosis	0.01293668854780638
amperage__skewness	0.011966365295760912
voltage__fft_aggregated__aggtype_centroid	0.011917213477507871
amperage__fft_coefficient__attr_real__coeff_0	0.011806265703240311
voltage__ratio_beyond_r_sigma__r_1	0.011798945198129598
amperage__count_above_mean	0.011473580355245238
voltage__cid_ce__normalize_True	0.010984462661446039
amperage__maximum	0.010248774233649121
voltage__spkt_welch_density__coeff_8	0.009944970279376863
voltage__ratio_beyond_r_sigma__r_3	0.009537525658278168
voltage__spkt_welch_density__coeff_5	0.009331617990929394
voltage__spkt_welch_density__coeff_2	0.008149999847418397
amperage__binned_entropy__max_bins_10	0.007932085128365279
voltage__ratio_beyond_r_sigma__r_1.5	0.007827687289100313

Suite à la page suivante

Tableau A.1 – suite

Caractéristique	Importance
amperage__ratio_beyond_r_sigma__r_1.5	0.007719005721871791
voltage__kurtosis	0.007582346049827874
amperage__spkt_welch_density__coeff_8	0.007412552508370637
amperage__friedrich_coefficients__coeff_1__m_3__r_30	0.007244728393547675
amperage__friedrich_coefficients__coeff_2__m_3__r_30	0.00706154320514575
voltage__skewness	0.007010174701365568
voltage__ratio_beyond_r_sigma__r_2	0.006763058175900595
amperage__ratio_beyond_r_sigma__r_1	0.006513814239108322
voltage__autocorrelation__lag_3	0.006482658508339959
amperage__ratio_beyond_r_sigma__r_2	0.006437455548227008
voltage__ratio_beyond_r_sigma__r_0.5	0.006313250253105528
voltage__friedrich_coefficients__coeff_1__m_3__r_30	0.006299274327185186
voltage__autocorrelation__lag_4	0.006068754720986213
voltage__minimum	0.005795914253937765
voltage__autocorrelation__lag_2	0.005722300338909823
voltage__friedrich_coefficients__coeff_0__m_3__r_30	0.00562201760155391
voltage__friedrich_coefficients__coeff_2__m_3__r_30	0.00546040435533202
voltage__maximum	0.005297536705755837
voltage__binned_entropy__max_bins_10	0.004621142601442257
amperage__fft_aggregated__aggtype_centroid	0.00437010327986771
amperage__minimum	0.004256772982944616
amperage__longest_strike_below_mean	0.003940453988653761
voltage__longest_strike_below_mean	0.003823057290337389

Suite à la page suivante

Tableau A.1 – suite

Caractéristique	Importance
amperage__index_mass_quantile__q_0.1	0.003558199269648321
voltage__longest_strike_above_mean	0.0032615039395746865
amperage__index_mass_quantile__q_0.7	0.003238672629328297
amperage__fft_coefficient__attr_real__coeff_1	0.003218228569315279
amperage__mean_second_derivative_central	0.0032006435205479676
amperage__fft_aggregated__aggtype_variance	0.0031512371330923066
amperage__index_mass_quantile__q_0.8	0.0031475429492972547
voltage__index_mass_quantile__q_0.3	0.003139483688036836
amperage__index_mass_quantile__q_0.3	0.0031300530967857516
voltage__index_mass_quantile__q_0.6	0.0030461146499057407
voltage__first_location_of_maximum	0.003003608231823569
voltage__index_mass_quantile__q_0.8	0.002986626248903004
amperage__index_mass_quantile__q_0.4	0.0029643228055359593
voltage__mean_second_derivative_central	0.002929083505552842
amperage__index_mass_quantile__q_0.9	0.0028524922448071232
voltage__index_mass_quantile__q_0.4	0.0028416272941684997
amperage__last_location_of_maximum	0.0028411245391078916
voltage__index_mass_quantile__q_0.1	0.0028336221500779653
voltage__fft_coefficient__attr_real__coeff_1	0.002830886621111853
amperage__index_mass_quantile__q_0.6	0.002809584456267076
amperage__longest_strike_above_mean	0.002797980827095108
voltage__first_location_of_minimum	0.0027893294239899834
voltage__time_reversal_asymmetry_statistic__lag_2	0.002776731675767426

Suite à la page suivante

Tableau A.1 – suite

Caractéristique	Importance
amperage__mean_change	0.0027540618769000195
amperage__first_location_of_minimum	0.0027514257673640617
amperage__index_mass_quantile__q_0.2	0.002734240259464592
voltage__index_mass_quantile__q_0.7	0.002733699428429342
voltage__last_location_of_minimum	0.002726641279349337
voltage__last_location_of_maximum	0.002703943713907172
amperage__first_location_of_maximum	0.0026703289065684145
voltage__index_mass_quantile__q_0.9	0.0026613261560901164
amperage__ratio_beyond_r_sigma__r_2.5	0.0026371374386773504
voltage__mean_change	0.0026299806262774466
voltage__time_reversal_asymmetry_statistic__lag_1	0.002581448759503157
amperage__time_reversal_asymmetry_statistic__lag_3	0.002550363683326409
voltage__time_reversal_asymmetry_statistic__lag_3	0.002545096353224607
amperage__time_reversal_asymmetry_statistic__lag_2	0.0025042365248565653
amperage__last_location_of_minimum	0.002446318230124937
voltage__index_mass_quantile__q_0.2	0.0024356486787782184
amperage__time_reversal_asymmetry_statistic__lag_1	0.0024091364967293604
voltage__large_standard_deviation__r_0.15	0.000678914089886819
voltage__large_standard_deviation__r_0.1	0.0006075372170943501
amperage__friedrich_coefficients__coeff_0__m_3__r_30	0.0003290173790657904
amperage__has_duplicate_max	0.00012321218065633992
amperage__has_duplicate_min	0.000053011857569736805
voltage__has_duplicate_min	0.00002233447316721565

Suite à la page suivante

Tableau A.1 – suite

Caractéristique	Importance
voltage__has_duplicate_max	0.00001573482535310156 3
amperage__length	0.0
amperage__variance_larger_than_standard_deviation	0.0
voltage__large_standard_deviation__r_0.05	0.0
amperage__large_standard_deviation__r_0.1	0.0
amperage__ratio_beyond_r_sigma__r_3	0.0
amperage__large_standard_deviation__r_0.15	0.0
amperage__large_standard_deviation__r_0.05	0.0
voltage__variance_larger_than_standard_deviation	0.0
voltage__length	0.0

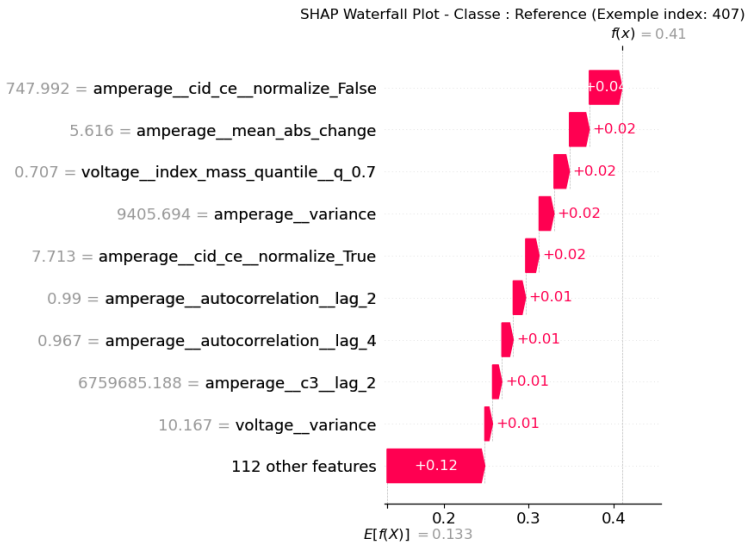


FIGURE A.1 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Référence » par un RF entraîné sur les données électriques.

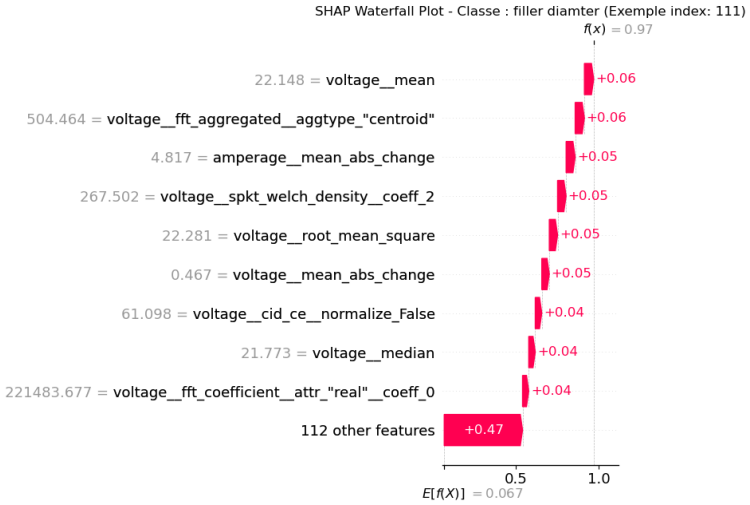


FIGURE A.2 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Diamètre fil d'apport » par un RF entraîné sur les données électriques.

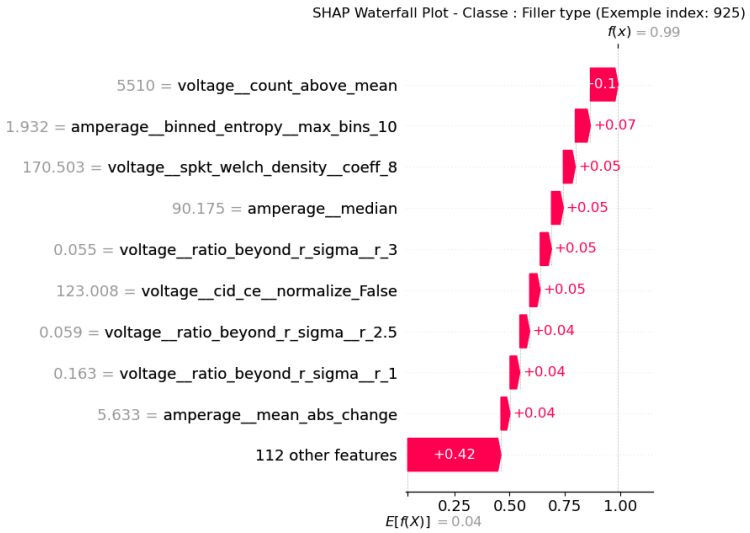


FIGURE A.3 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Type de fil d'apport » par un RF entraîné sur les données électriques.

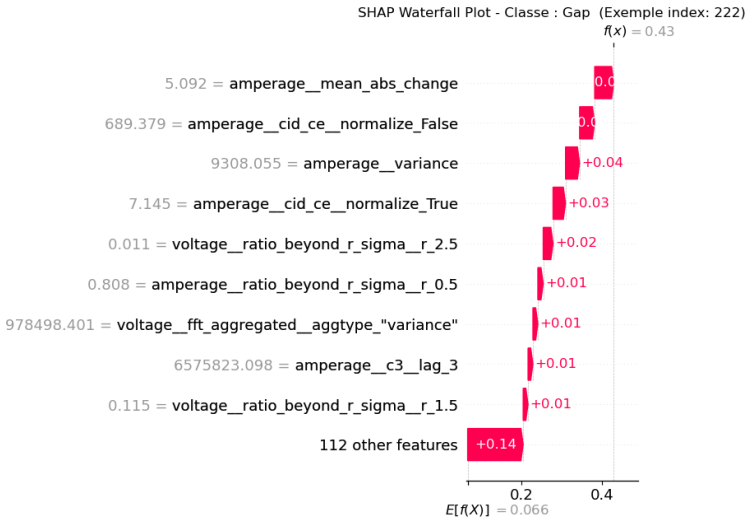


FIGURE A.4 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Écartement » par un RF entraîné sur les données électriques.

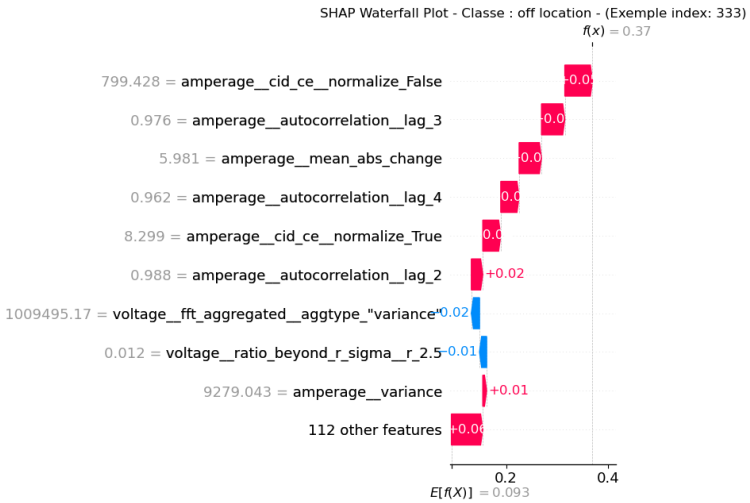


FIGURE A.5 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Hors-position - » par un RF entraîné sur les données électriques.

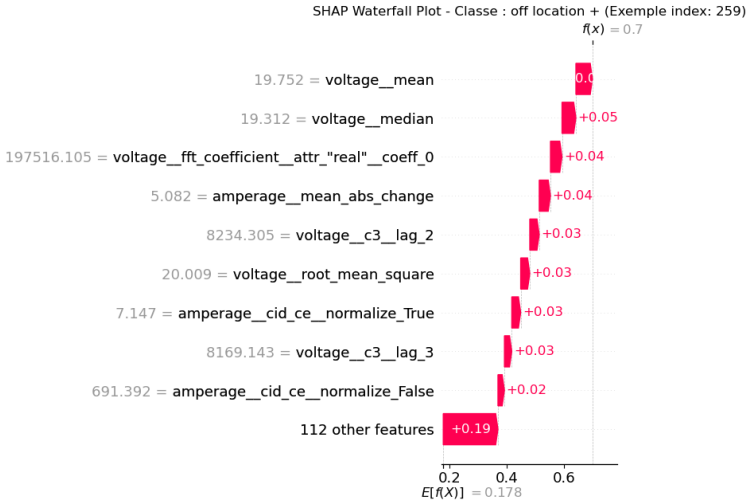


FIGURE A.6 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Hors-position + » par un RF entraîné sur les données électriques.

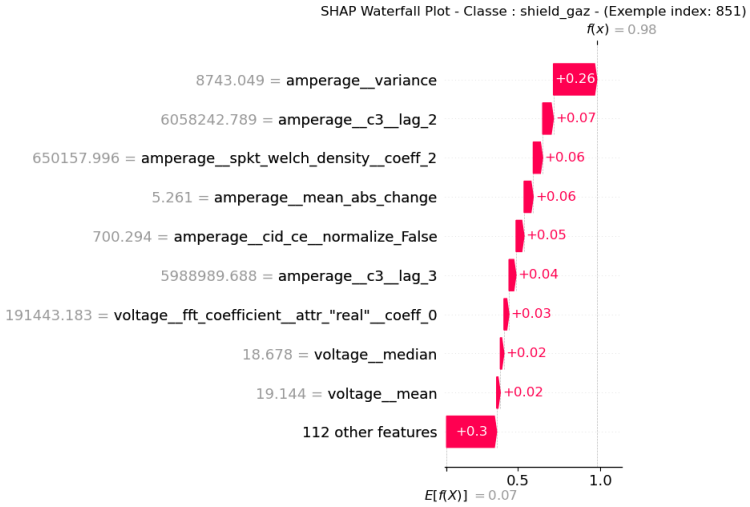


FIGURE A.7 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Débit de gaz - » par un RF entraîné sur les données électriques.

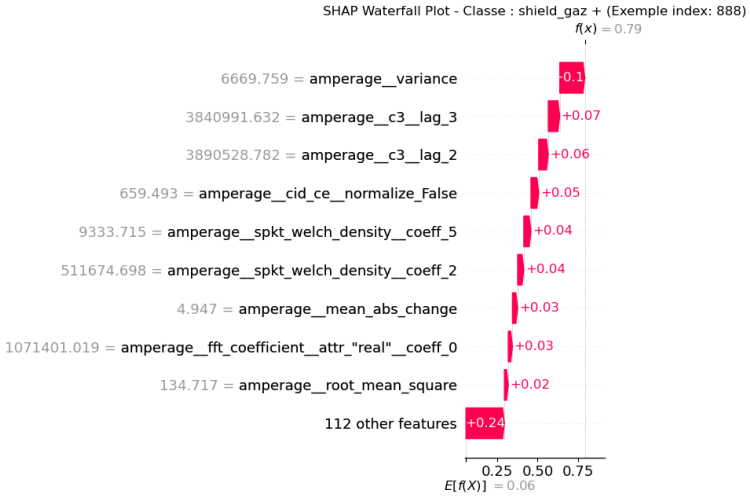


FIGURE A.8 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Débit de gaz + » par un RF entraîné sur les données électriques.

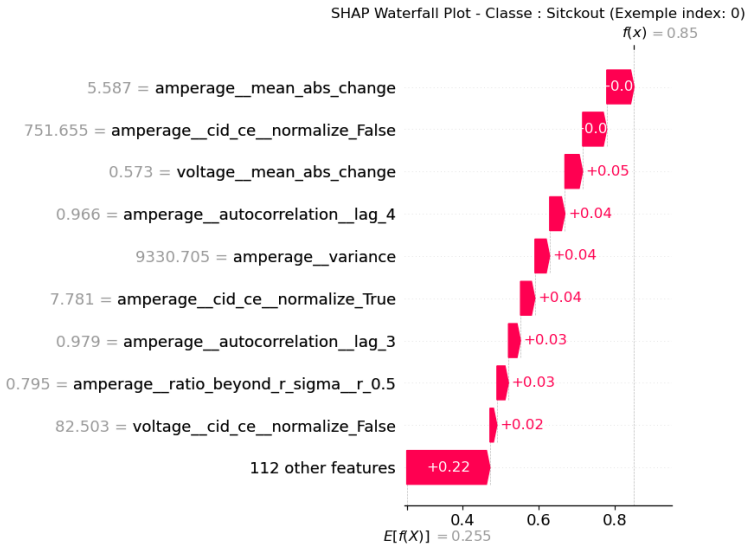


FIGURE A.9 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Longueur d'arc » par un RF entraîné sur les données électriques.

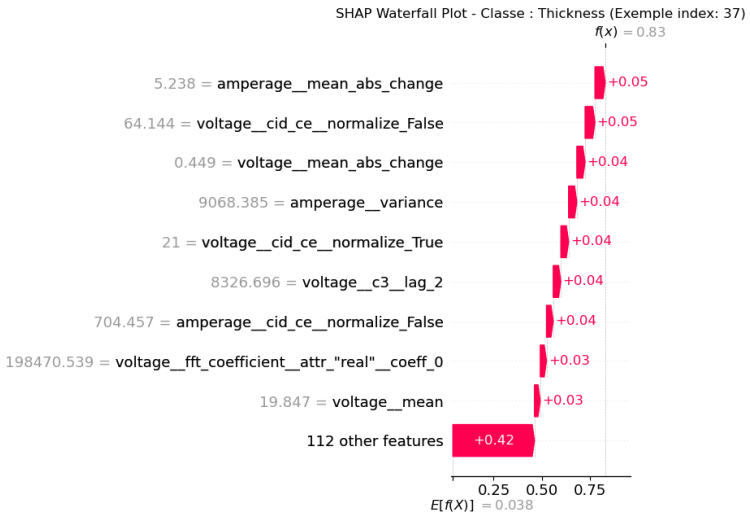


FIGURE A.10 : Exemple d'importance des caractéristiques calculée par SHAP pour la prédiction de la classe « Épaisseur » par un RF entraîné sur les données électriques.